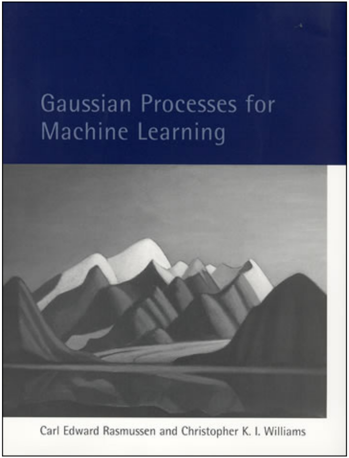
Session 1: Gaussian Processes

Neil D. Lawrence

BioPreDyn Workshop
Barcelona, 12th June 2012

Outline

Book


Gaussian Processes for
Machine Learning
Carl Edward Rasmussen and Christopher K. I. Williams

?

Outline

1. The Gaussian Density
2. GP Limitations
3. Gene Expression Examples
4. Conclusions

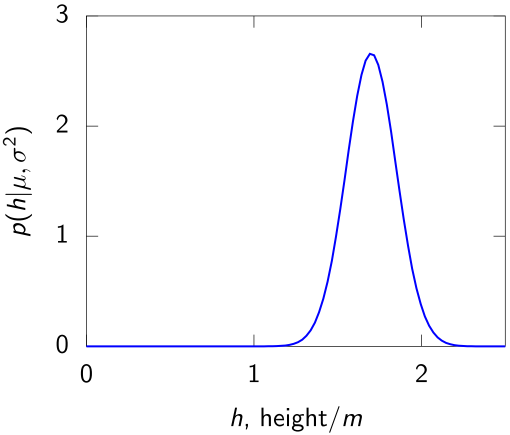
The Gaussian Density

- Perhaps the most common probability density.

$$p(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$
$$= \mathcal{N}\left(y|\mu, \sigma^2\right)$$

- The Gaussian density.

Gaussian Density



The Gaussian PDF with $\mu = 1.7$ and variance $\sigma^2 = 0.0225$. Mean shown as red line. It could represent the heights of a population of students.

Gaussian Density

$$\mathcal{N}\left(y|\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

Two Important Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

(*Aside*: As sum increases, sum of non-Gaussian, finite variance variables is also Gaussian [central limit theorem].)

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Two Important Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

(*Aside*: As sum increases, sum of non-Gaussian, finite variance variables is also Gaussian [central limit theorem].)

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Two Important Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

(*Aside*: As sum increases, sum of non-Gaussian, finite variance variables is also Gaussian [central limit theorem].)

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Two Important Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

(*Aside*: As sum increases, sum of non-Gaussian, finite variance variables is also Gaussian [central limit theorem].)

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Two Important Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

(*Aside*: As sum increases, sum of non-Gaussian, finite variance variables is also Gaussian [central limit theorem].)

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Two Important Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

(*Aside*: As sum increases, sum of non-Gaussian, finite variance variables is also Gaussian [central limit theorem].)

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Two Simultaneous Equations

A system of two differential equations with two unknowns.

$$
\begin{aligned}
y_1 &= mt_1 + c \\
y_2 &= mt_2 + c
\end{aligned}
$$

Two Simultaneous Equations

A system of two differential equations with two unknowns.

$$y_1 - y_2 = m(t_1 - t_2)$$

Two Simultaneous Equations

A system of two differential equations with two unknowns.

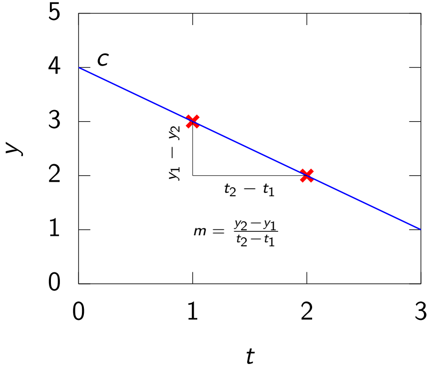
$$\frac{y_1 - y_2}{t_1 - t_2} = m$$

Two Simultaneous Equations

A system of two differential equations with two unknowns.

$$m = \frac{y_2 - y_1}{t_2 - t_1}$$

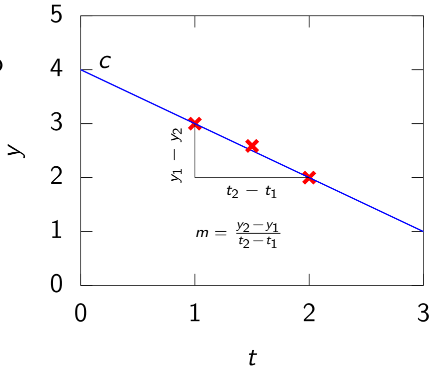
$$c = y_1 - mt_1$$



Two Simultaneous Equations

How do we deal with three simultaneous equations with only two unknowns?

$$y_1 = mt_1 + c$$
$$y_2 = mt_2 + c$$
$$y_3 = mt_3 + c$$



Overdetermined System

- With two unknowns and two observations:

$$\begin{aligned} y_1 =& mt_1 + c \\ y_2 =& mt_2 + c \end{aligned}$$

- Additional observation leads to *overdetermined* system.

$$y_3 = mt_3 + c$$

- This problem is solved through a noise model $\epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$

$$\begin{aligned} y_1 &= mt_1 + c + \epsilon_1 \\ y_2 &= mt_2 + c + \epsilon_2 \\ y_3 &= mt_3 + c + \epsilon_3 \end{aligned}$$

Overdetermined System

- With two unknowns and two observations:

$$y_1 = mt_1 + c$$
$$y_2 = mt_2 + c$$

- Additional observation leads to *overdetermined* system.

$$y_3 = mt_3 + c$$

- This problem is solved through a noise model $\epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$

$$y_1 = mt_1 + c + \epsilon_1$$
$$y_2 = mt_2 + c + \epsilon_2$$
$$y_3 = mt_3 + c + \epsilon_3$$

Overdetermined System

- With two unknowns and two observations:

$$y_1 = mt_1 + c$$
$$y_2 = mt_2 + c$$

- Additional observation leads to *overdetermined* system.

$$y_3 = mt_3 + c$$

- This problem is solved through a noise model $\epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$

$$y_1 = mt_1 + c + \epsilon_1$$
$$y_2 = mt_2 + c + \epsilon_2$$
$$y_3 = mt_3 + c + \epsilon_3$$

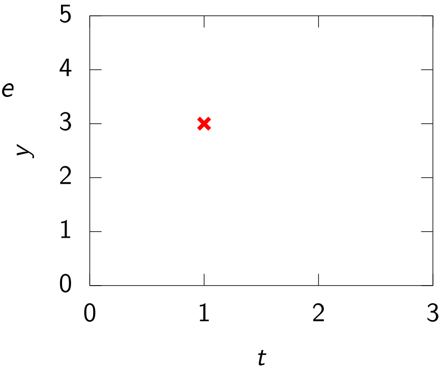
Noise Models

- We aren't modeling entire system.
- Noise model gives mismatch between model and data.
- Gaussian model justified by appeal to central limit theorem.
- Other models also possible (Student-t for heavy tails).
- Maximum likelihood with Gaussian noise leads to *least squares*.

Underdetermined System

What about two unknowns and *one* observation?

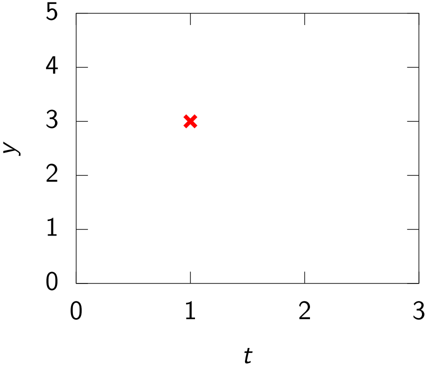
$$y_1 = mt_1 + c$$



Underdetermined System

Can compute m given c.

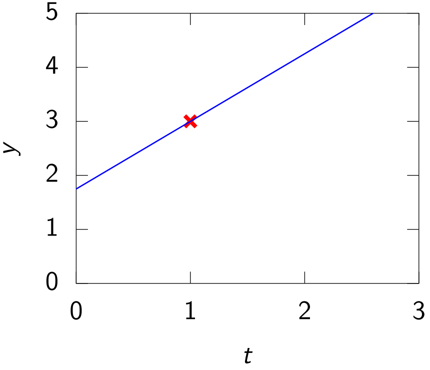
$$m = \frac{y_1 - c}{t}$$



Underdetermined System

Can compute m given c.

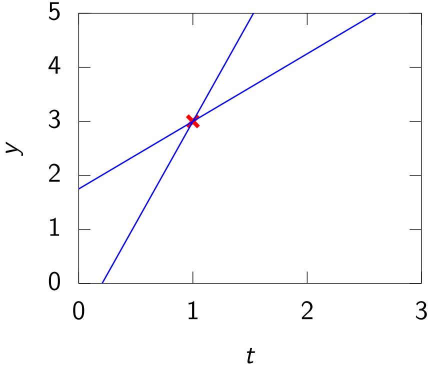
$$c = 1.75 \Longrightarrow m = 1.25$$



Underdetermined System

Can compute m given c.

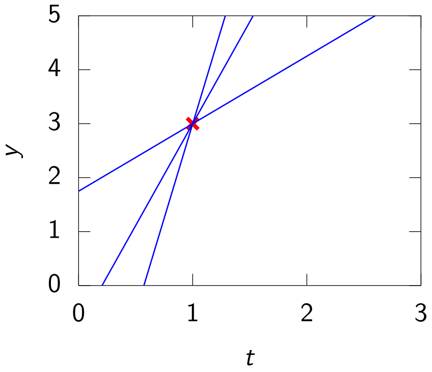
$$c = -0.777 \implies m = 3.78$$



Underdetermined System

Can compute m given c.

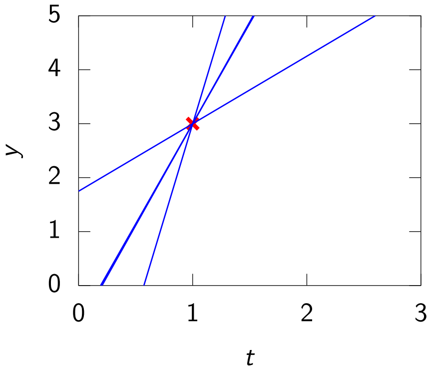
$$c = -4.01 \Longrightarrow m = 7.01$$



Underdetermined System

Can compute m given c.

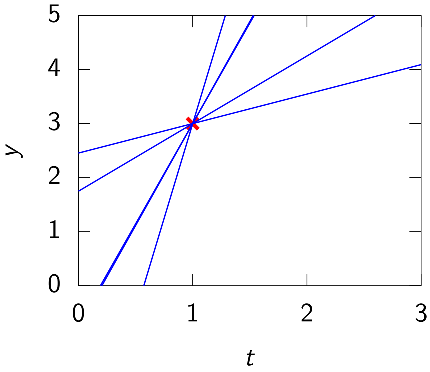
$$c = -0.718 \Longrightarrow m = 3.72$$



Underdetermined System

Can compute m given c.

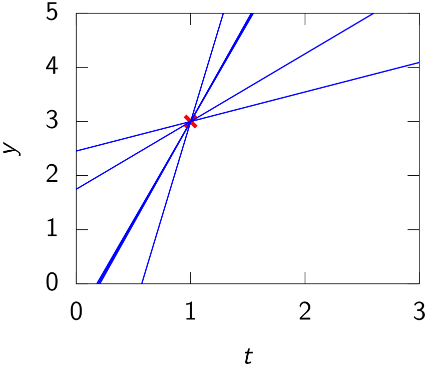
$$c = 2.45 \Longrightarrow m = 0.545$$



Underdetermined System

Can compute m given c.

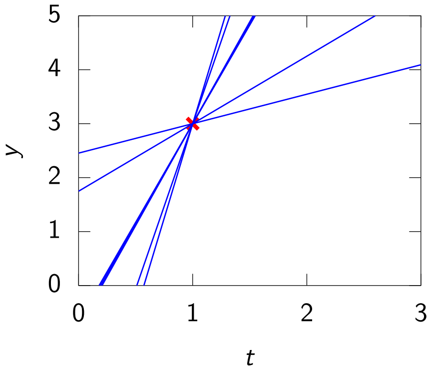
$$c = -0.657 \implies m = 3.66$$



Underdetermined System

Can compute *m* given *c*.

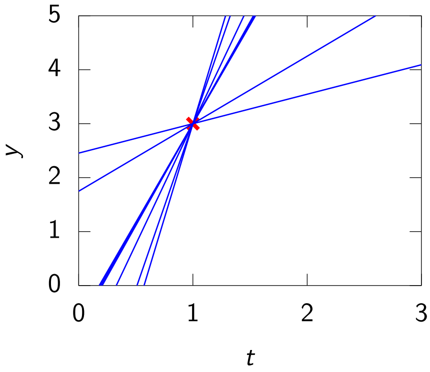
$$c = -3.13 \implies m = 6.13$$



Underdetermined System

Can compute m given c.

$$c = -1.47 \Longrightarrow m = 4.47$$

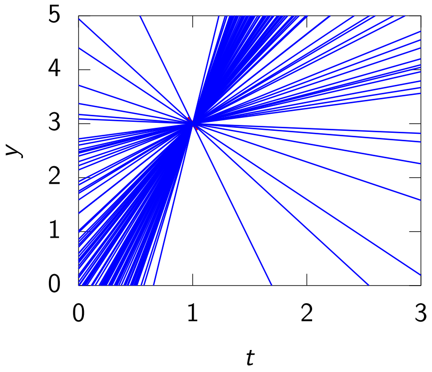


Underdetermined System

Can compute m given c.
Assume

$$c \sim \mathcal{N}(0, 4),$$

we find a distribution of solutions.



Probability for Under- and Overdetermined

- To deal with overdetermined introduced probability distribution for 'variable', ϵ_i.
- For underdetermined system introduced probability distribution for 'parameter', c.
- This is known as a Bayesian treatment.

- For general Bayesian inference need multivariate priors.
- E.g. for multivariate linear regression:

$$y_i = \sum_i w_j t_{i,j} + \epsilon_i$$

(where we've dropped c for convenience), we need a prior over $\mathbf{w}$.
- This motivates a *multivariate* Gaussian density.
- We will use the multivariate Gaussian to put a prior *directly* on the function (a Gaussian process).

- For general Bayesian inference need multivariate priors.
- E.g. for multivariate linear regression:

$$y_i = \mathbf{w}^\top \mathbf{t}_{i,:} + \epsilon_i$$

(where we've dropped c for convenience), we need a prior over $\mathbf{w}$.
- This motivates a *multivariate* Gaussian density.
- We will use the multivariate Gaussian to put a prior *directly* on the function (a Gaussian process).

Two Dimensional Gaussian

- Consider height, h/m and weight, w/kg.
- Could sample height from a distribution:

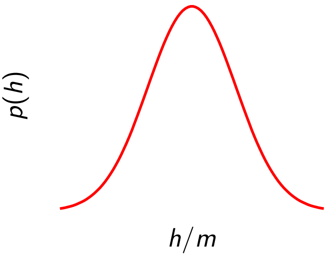
$$p(h) \sim \mathcal{N}(1.7, 0.0225)$$

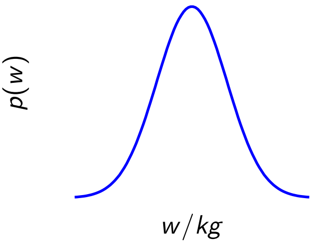
- And similarly weight:

$$p(w) \sim \mathcal{N}(75, 36)$$

Height and Weight Models

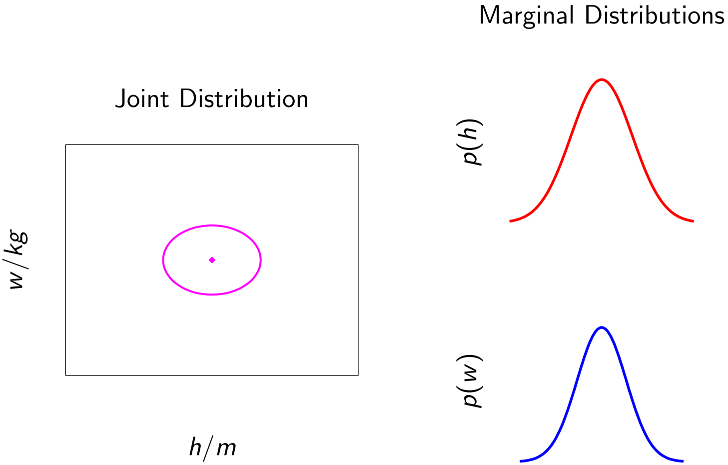
Marginal Distributions





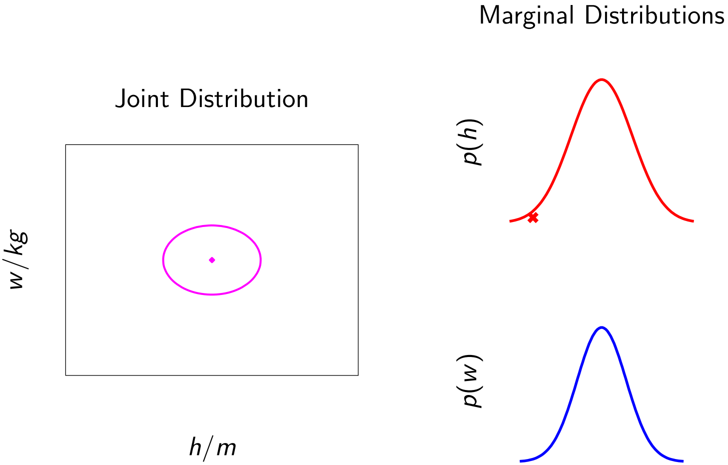
Gaussian distributions for height and weight.

Sampling Two Dimensional Variables



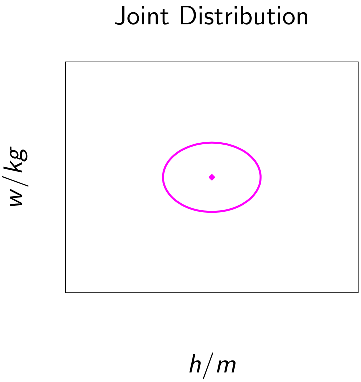
Sample height and weight one after the other and plot against each other.

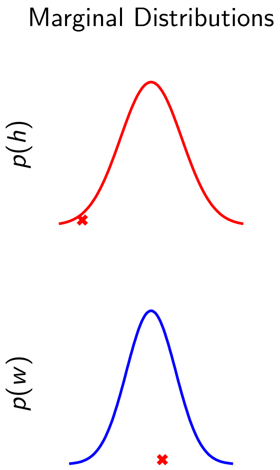
Sampling Two Dimensional Variables



Sample height and weight one after the other and plot against each other.

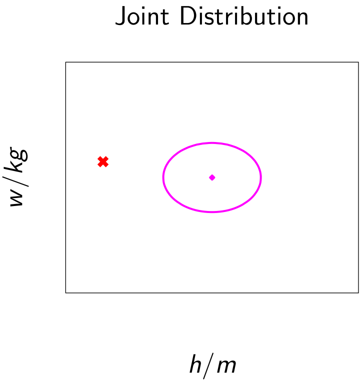
Sampling Two Dimensional Variables

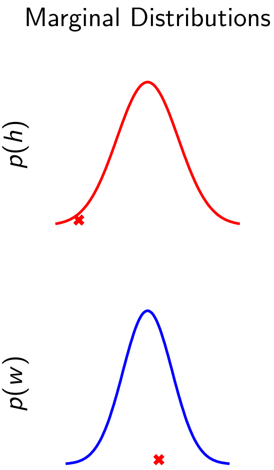




Sample height and weight one after the other and plot against each other.

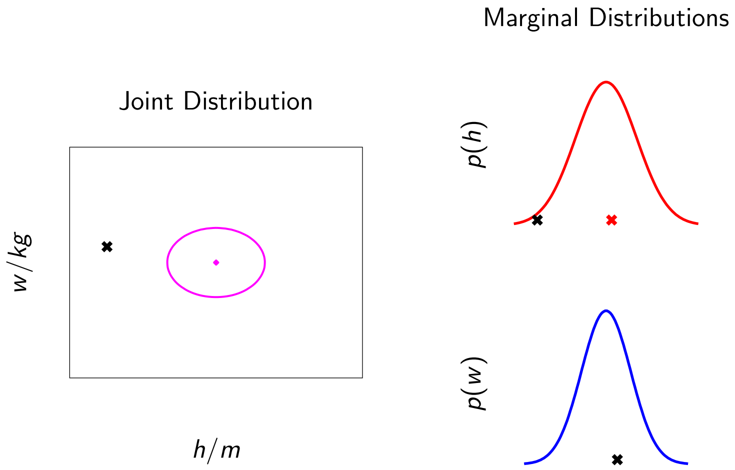
Sampling Two Dimensional Variables





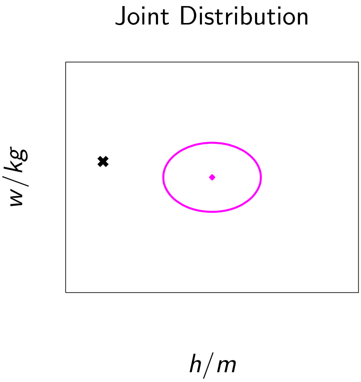
Sample height and weight one after the other and plot against each other.

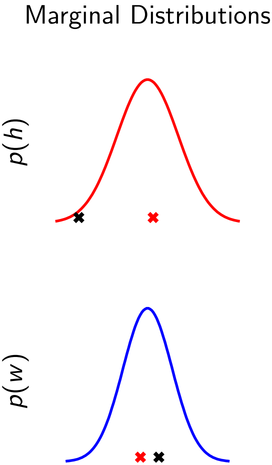
Sampling Two Dimensional Variables



Sample height and weight one after the other and plot against each other.

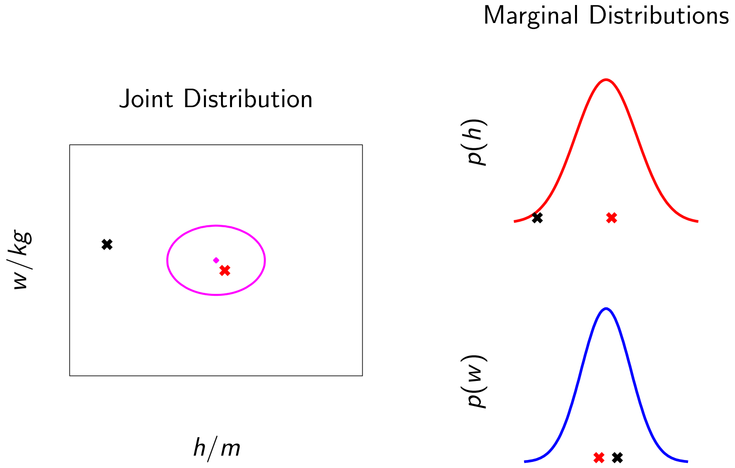
Sampling Two Dimensional Variables





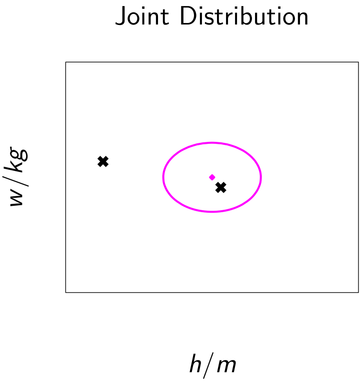
Sample height and weight one after the other and plot against each other.

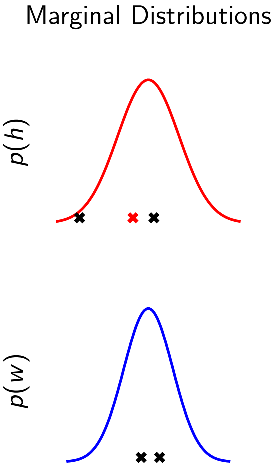
Sampling Two Dimensional Variables



Sample height and weight one after the other and plot against each other.

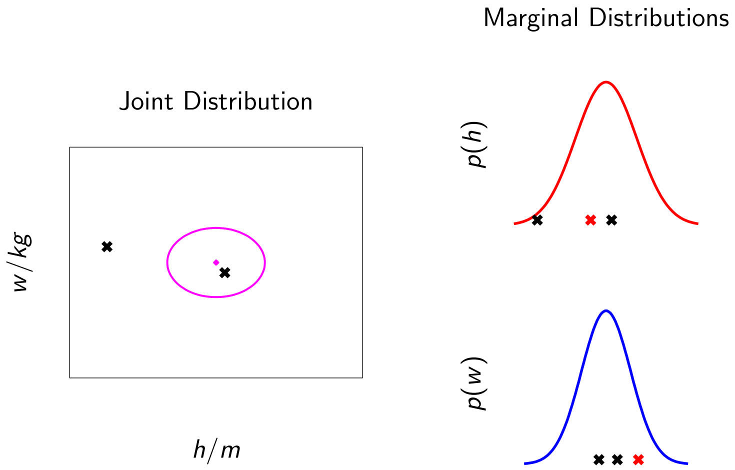
Sampling Two Dimensional Variables





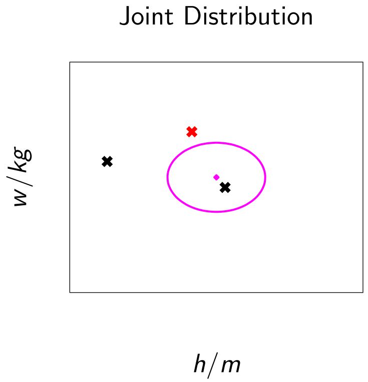
Sample height and weight one after the other and plot against each other.

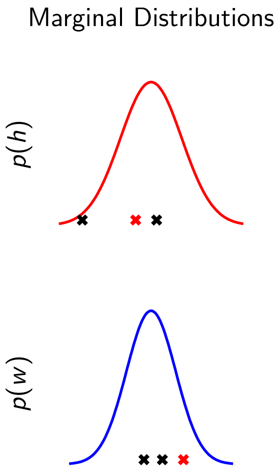
Sampling Two Dimensional Variables



Sample height and weight one after the other and plot against each other.

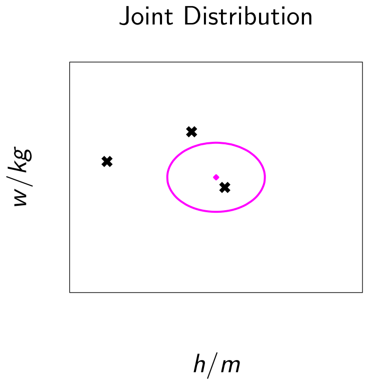
Sampling Two Dimensional Variables

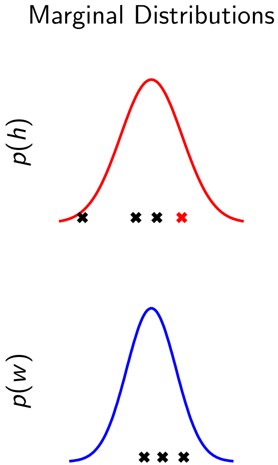




Sample height and weight one after the other and plot against each other.

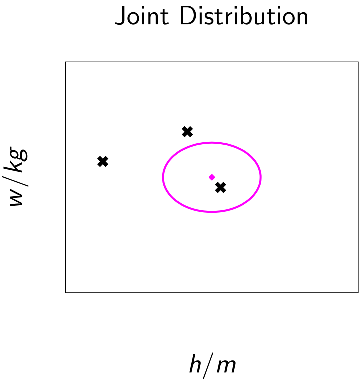
Sampling Two Dimensional Variables

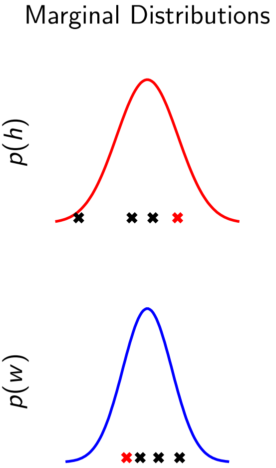




Sample height and weight one after the other and plot against each other.

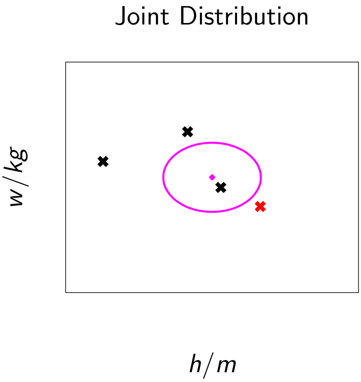
Sampling Two Dimensional Variables

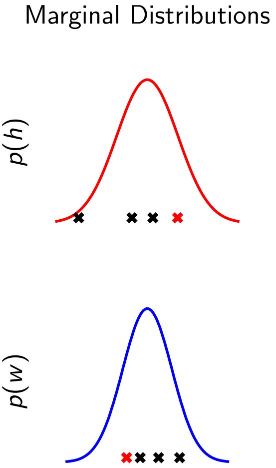




Sample height and weight one after the other and plot against each other.

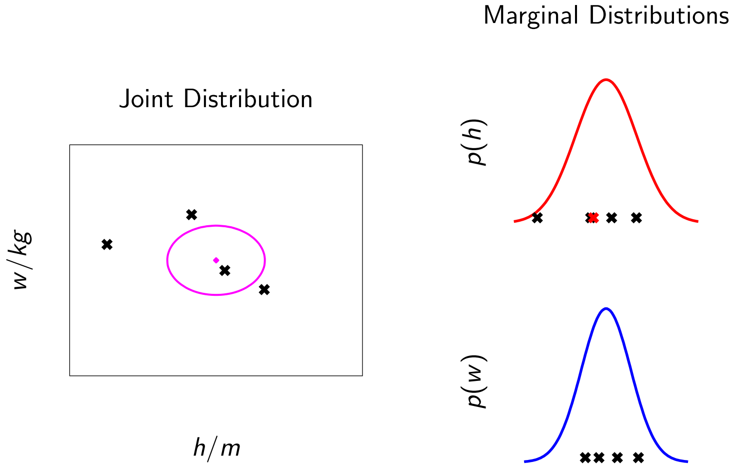
Sampling Two Dimensional Variables





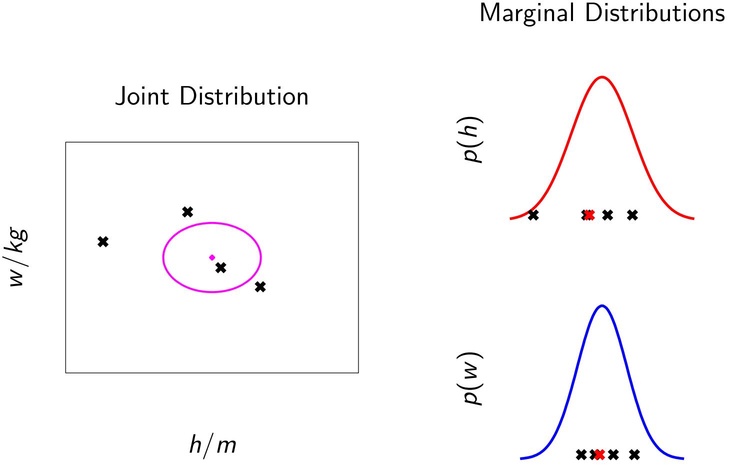
Sample height and weight one after the other and plot against each other.

Sampling Two Dimensional Variables



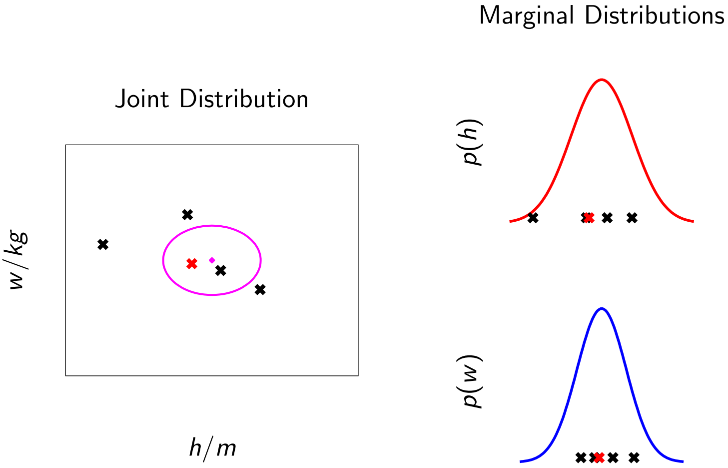
Sample height and weight one after the other and plot against each other.

Sampling Two Dimensional Variables



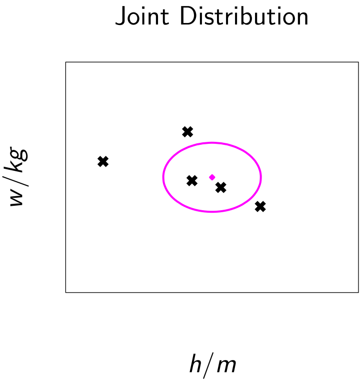
Sample height and weight one after the other and plot against each other.

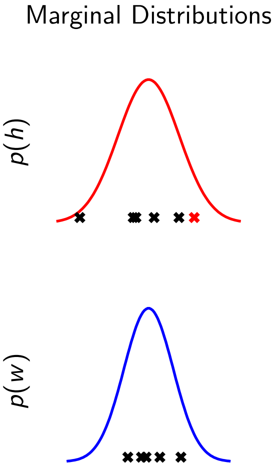
Sampling Two Dimensional Variables



Sample height and weight one after the other and plot against each other.

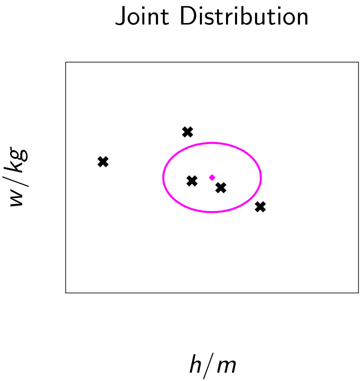
Sampling Two Dimensional Variables

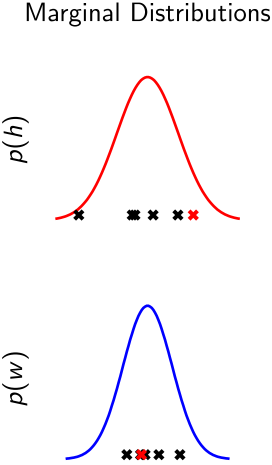




Sample height and weight one after the other and plot against each other.

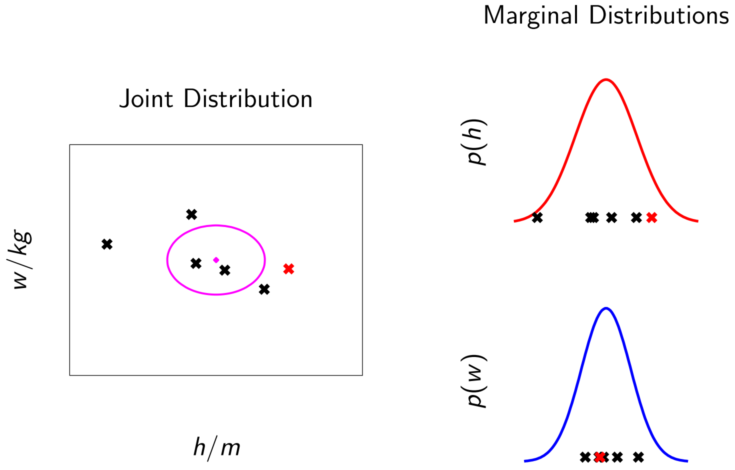
Sampling Two Dimensional Variables





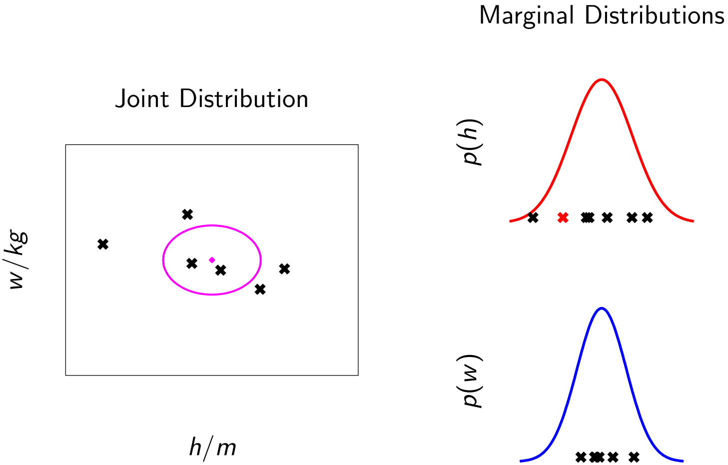
Sample height and weight one after the other and plot against each other.

Sampling Two Dimensional Variables



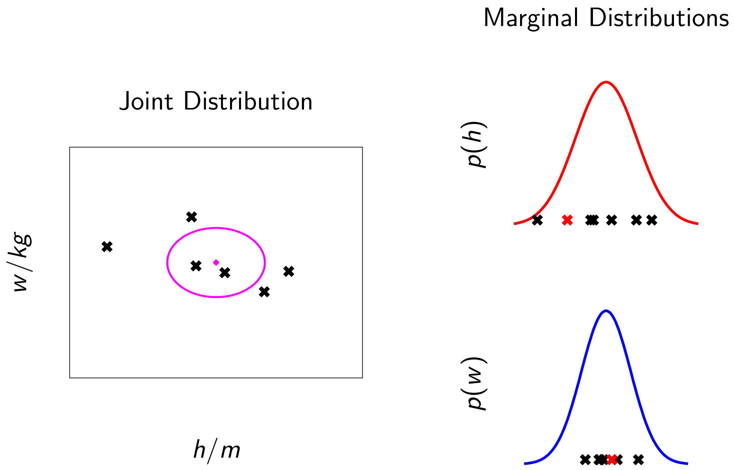
Sample height and weight one after the other and plot against each other.

Sampling Two Dimensional Variables



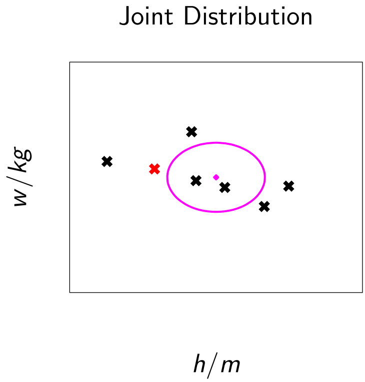
Sample height and weight one after the other and plot against each other.

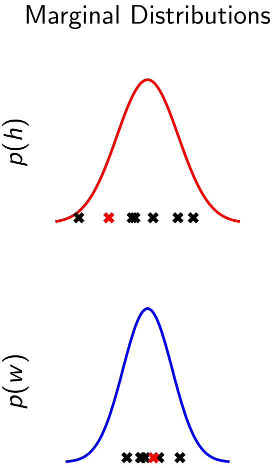
Sampling Two Dimensional Variables



Sample height and weight one after the other and plot against each other.

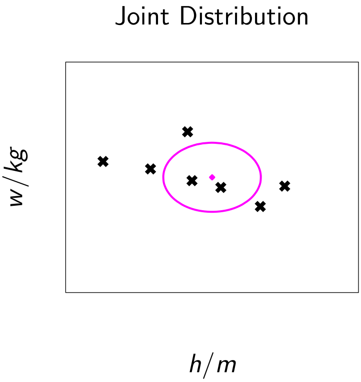
Sampling Two Dimensional Variables

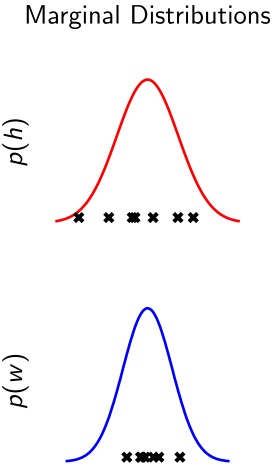




Sample height and weight one after the other and plot against each other.

Sampling Two Dimensional Variables





Sample height and weight one after the other and plot against each other.

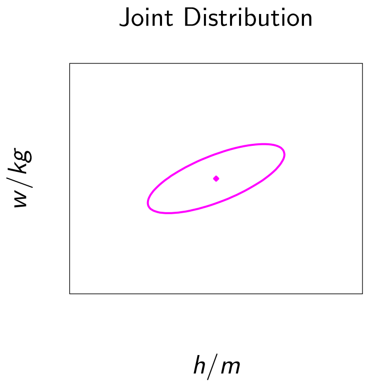
Independence Assumption

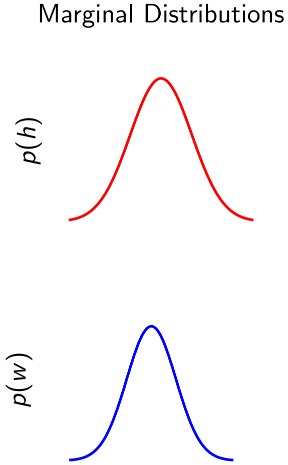
- This assumes height and weight are independent.

$$p(h, w) = p(h)p(w)$$

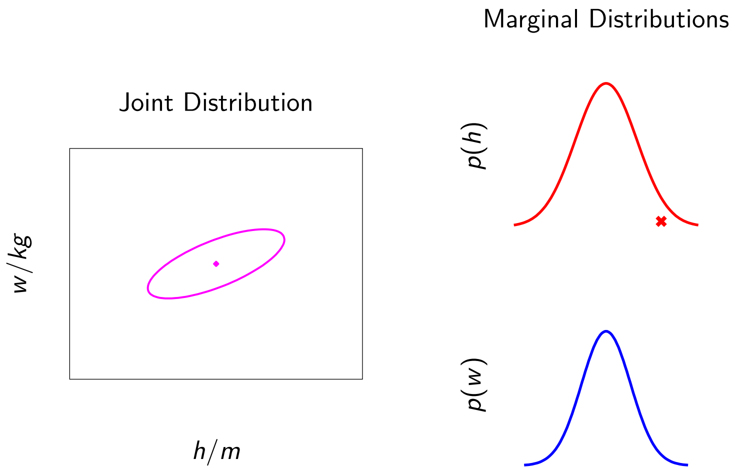
- In reality they are dependent (body mass index) $= \frac{w}{h^2}$.

Sampling Two Dimensional Variables

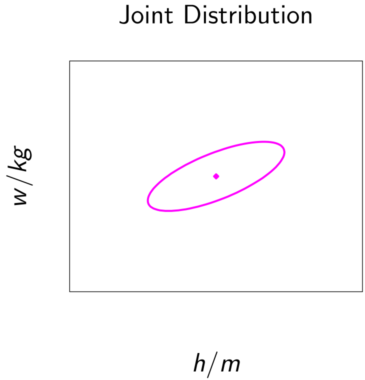


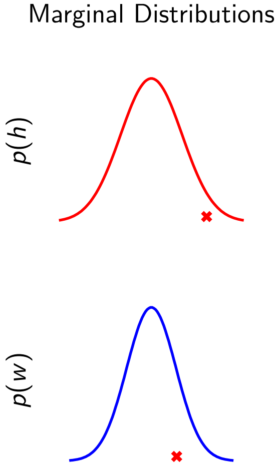


Sampling Two Dimensional Variables

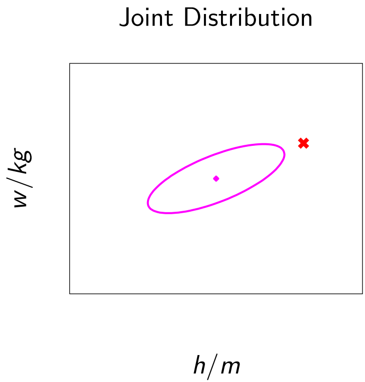


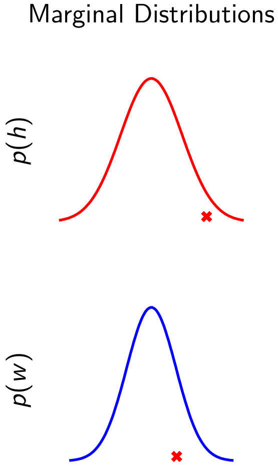
Sampling Two Dimensional Variables



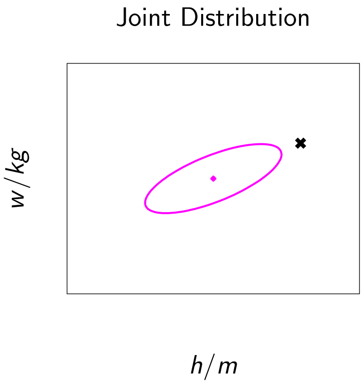


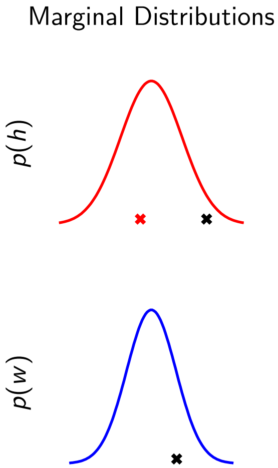
Sampling Two Dimensional Variables



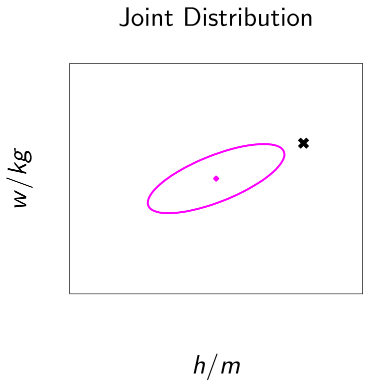


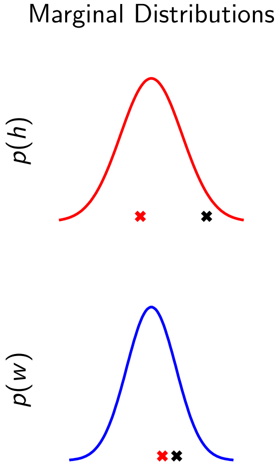
Sampling Two Dimensional Variables



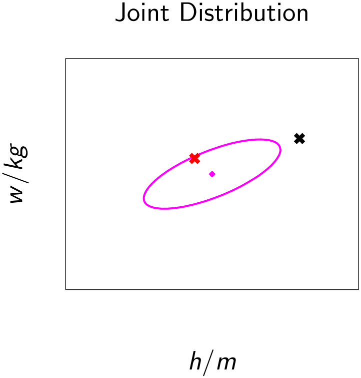


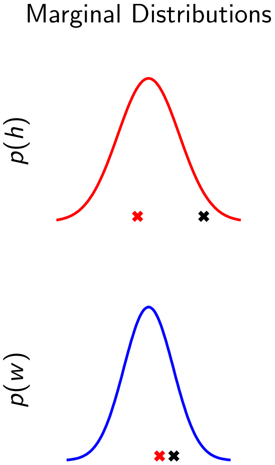
Sampling Two Dimensional Variables



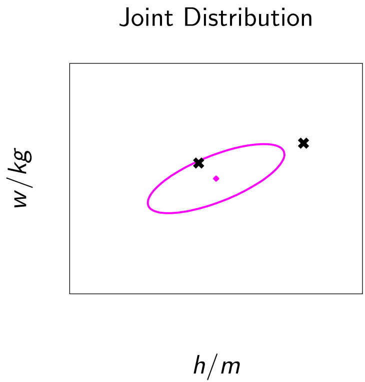


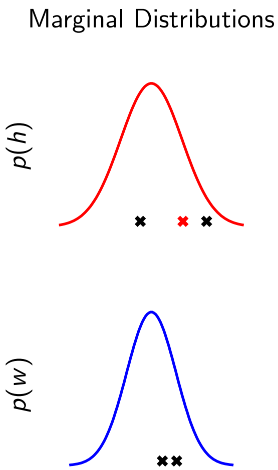
Sampling Two Dimensional Variables



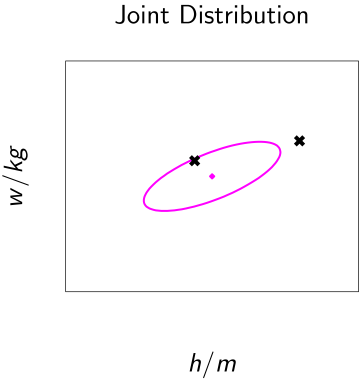


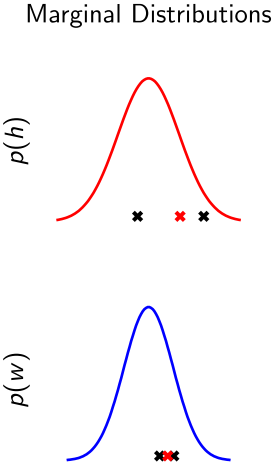
Sampling Two Dimensional Variables



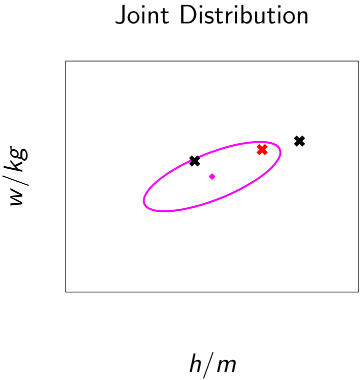


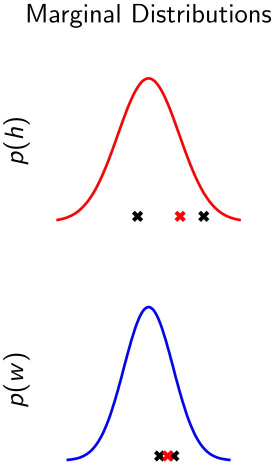
Sampling Two Dimensional Variables



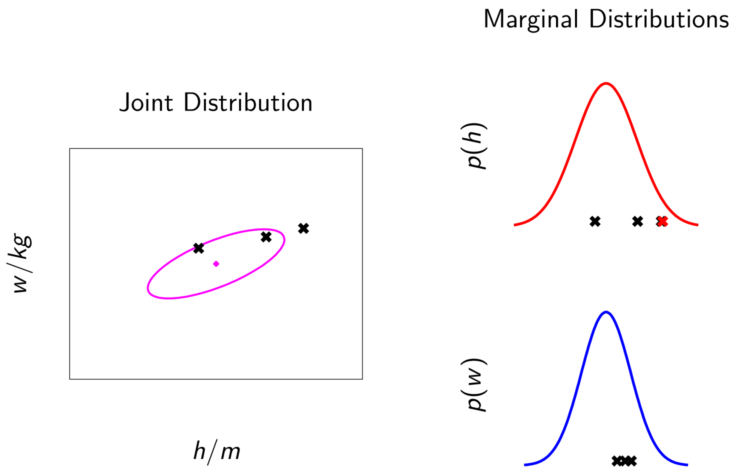


Sampling Two Dimensional Variables

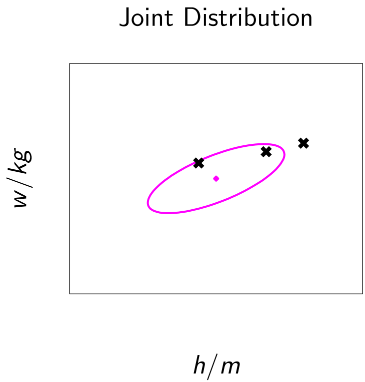


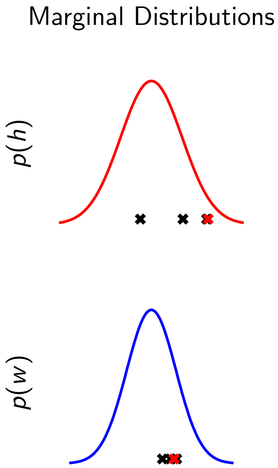


Sampling Two Dimensional Variables

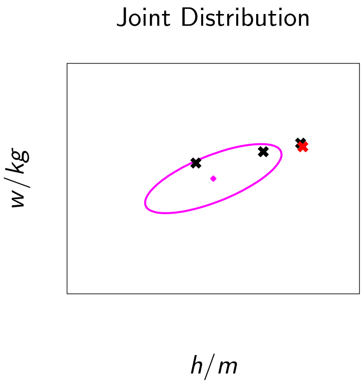


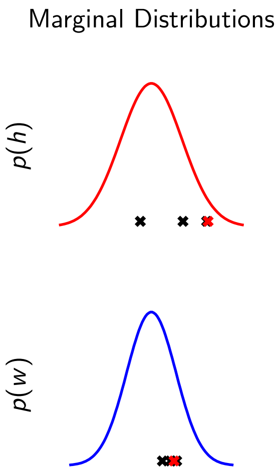
Sampling Two Dimensional Variables



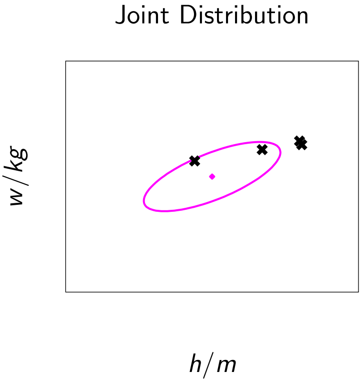


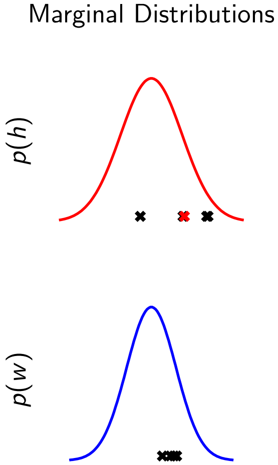
Sampling Two Dimensional Variables



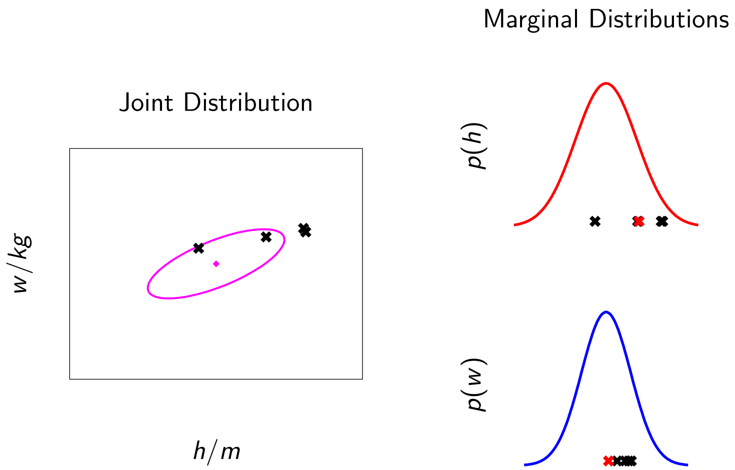


Sampling Two Dimensional Variables

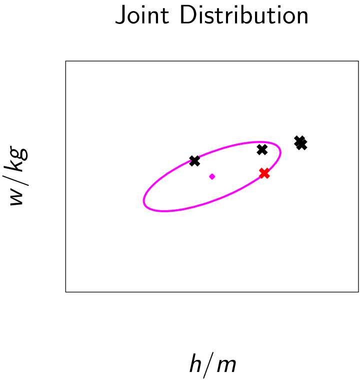


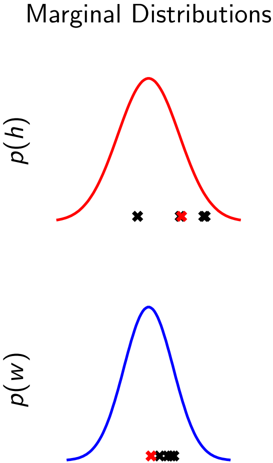


Sampling Two Dimensional Variables

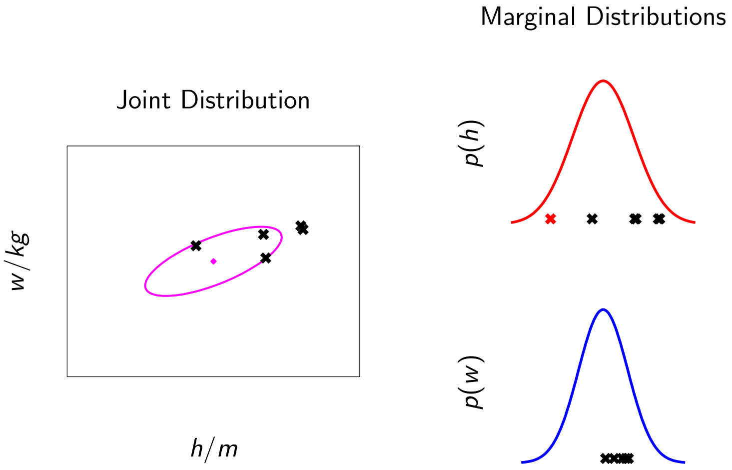


Sampling Two Dimensional Variables

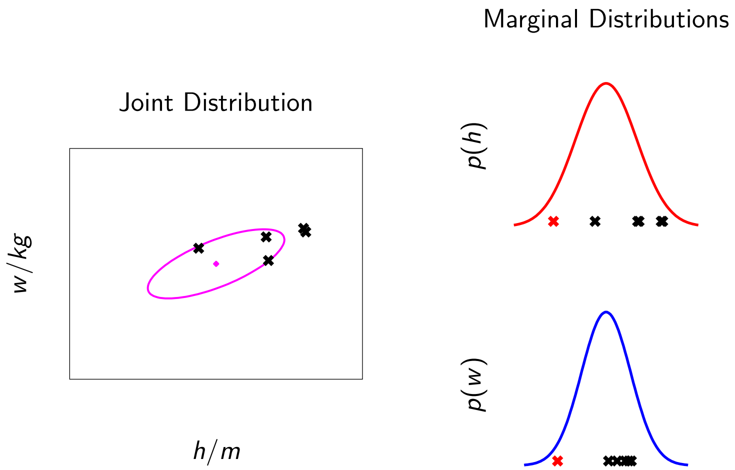




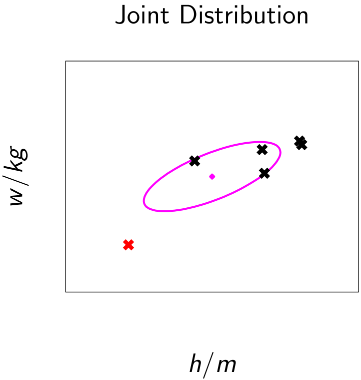
Sampling Two Dimensional Variables

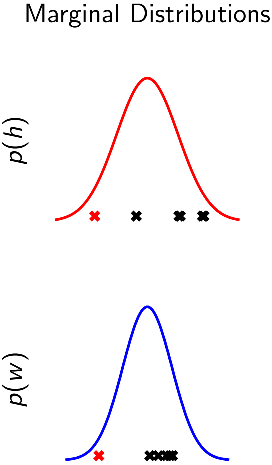


Sampling Two Dimensional Variables

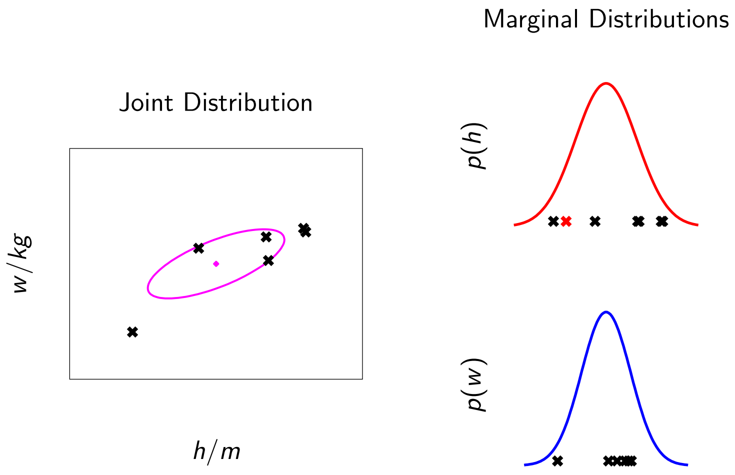


Sampling Two Dimensional Variables

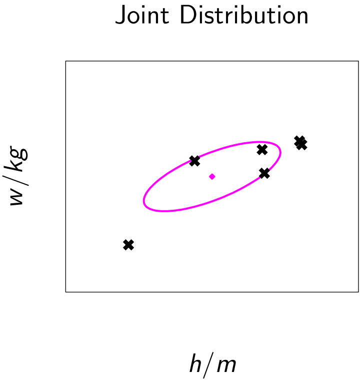


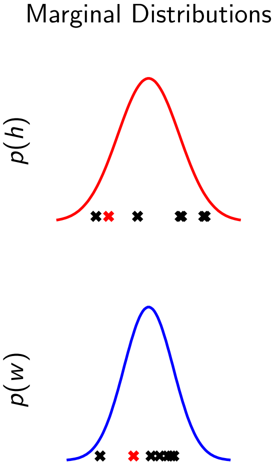


Sampling Two Dimensional Variables

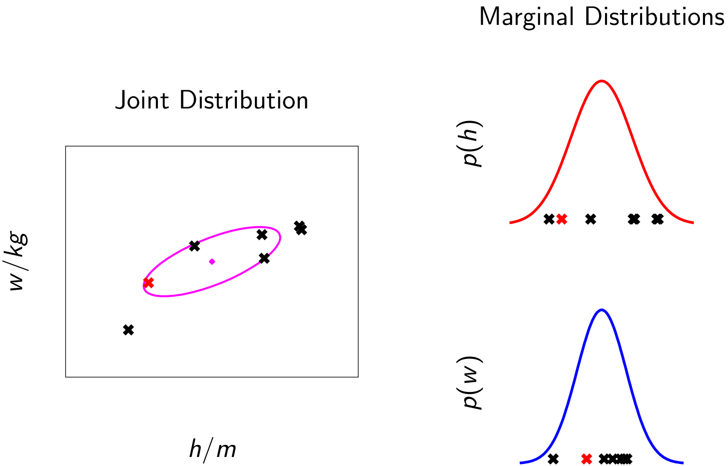


Sampling Two Dimensional Variables

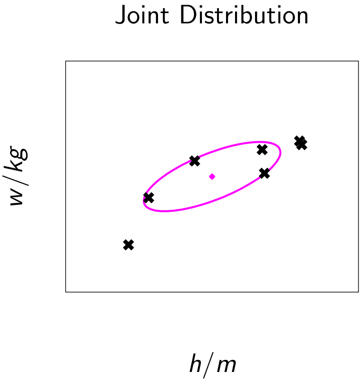


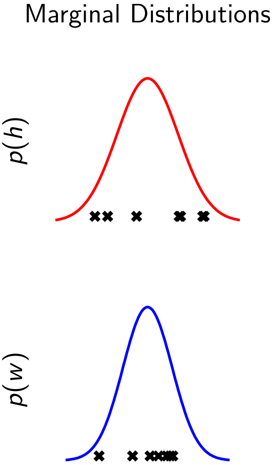


Sampling Two Dimensional Variables



Sampling Two Dimensional Variables





Independent Gaussians

$$p(w, h) = p(w)p(h)$$

Independent Gaussians

$$p(w, h) = \frac{1}{\sqrt{2\pi\sigma_1^2}\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\left(\frac{(w-\mu_1)^2}{\sigma_1^2} + \frac{(h-\mu_2)^2}{\sigma_2^2}\right)\right)$$

Independent Gaussians

$$p(w,h) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} w \\ h \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)^\top \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1} \left(\begin{bmatrix} w \\ h \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)\right)$$

Independent Gaussians

$$p(\mathbf{y}) = \frac{1}{2\pi \left|\mathbf{D}\right|} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\top}\mathbf{D}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)$$

Correlated Gaussian

Form correlated from original by rotating the data space using matrix **R**.

$$p(\mathbf{y}) = \frac{1}{2\pi \left|\mathbf{D}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^{\top}\mathbf{D}^{-1}(\mathbf{y} - \boldsymbol{\mu})\right)$$

Correlated Gaussian

Form correlated from original by rotating the data space using matrix **R**.

$$p(\mathbf{y}) = \frac{1}{2\pi \left|\mathbf{D}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{R}^\top\mathbf{y} - \mathbf{R}^\top\boldsymbol{\mu})^\top\mathbf{D}^{-1}(\mathbf{R}^\top\mathbf{y} - \mathbf{R}^\top\boldsymbol{\mu})\right)$$

Correlated Gaussian

Form correlated from original by rotating the data space using matrix **R**.

$$p(\mathbf{y}) = \frac{1}{2\pi \left|\mathbf{D}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \mathbf{R}\mathbf{D}^{-1}\mathbf{R}^\top(\mathbf{y}-\boldsymbol{\mu})\right)$$

this gives a covariance matrix:

$$\mathbf{C}^{-1} = \mathbf{R}\mathbf{D}^{-1}\mathbf{R}^\top$$

Correlated Gaussian

Form correlated from original by rotating the data space using matrix **R**.

$$p(\mathbf{y}) = \frac{1}{2\pi \left|\mathbf{C}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\top}\mathbf{C}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)$$

this gives a covariance matrix:

$$\mathbf{C} = \mathbf{R}\mathbf{D}\mathbf{R}^{\top}$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Multivariate Consequence

- If

$$\mathbf{t} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- And

$$\mathbf{y} = \mathbf{W}\mathbf{t}$$

- Then

$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{W}\mu, \mathbf{W}\Sigma\mathbf{W}^{\top}\right)$$

Multivariate Consequence

- If

$$\mathbf{t} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- And

$$\mathbf{y} = \mathbf{W}\mathbf{t}$$

- Then

$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{W}\mu, \mathbf{W}\Sigma\mathbf{W}^\top\right)$$

Multivariate Consequence

- If

$$\mathbf{t} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- And

$$\mathbf{y} = \mathbf{W}\mathbf{t}$$

- Then

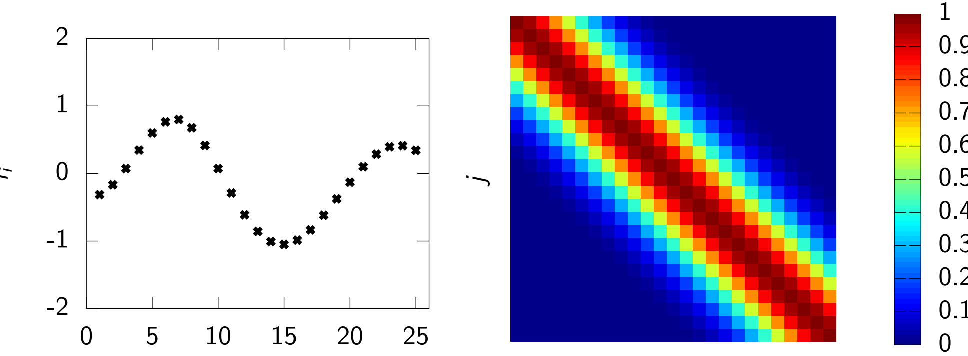
$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{W}\mu, \mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^{\top}\right)$$

Sampling a Function

Multi-variate Gaussians

- We will consider a Gaussian with a particular structure of covariance matrix.
- Generate a single sample from this 25 dimensional Gaussian distribution, $\mathbf{f} = [f_1, f_2 \ldots f_{25}]$.
- We will plot these points against their index.

Gaussian Distribution Sample

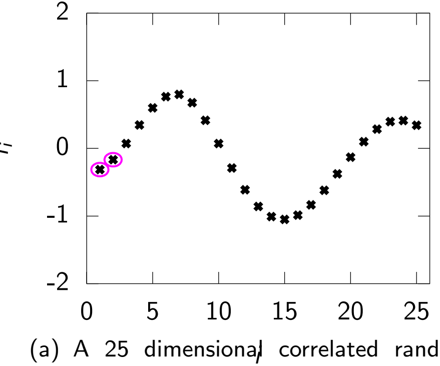


(a) A 25 dimensional correlated random variable (values ploted against index)

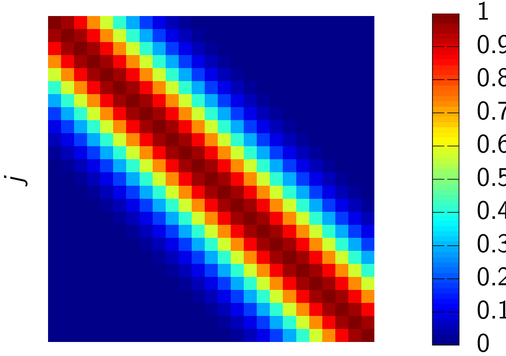
(b) colormap showing correlations between dimensions.

Figure: A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



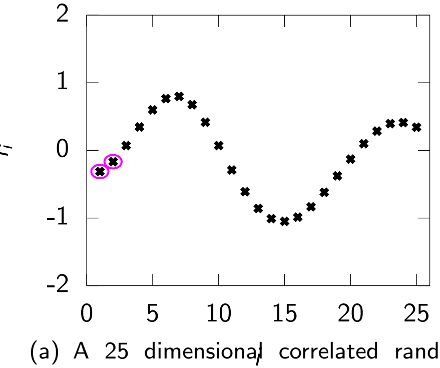
(a) A 25 dimensional correlated random variable (values ploted against index)



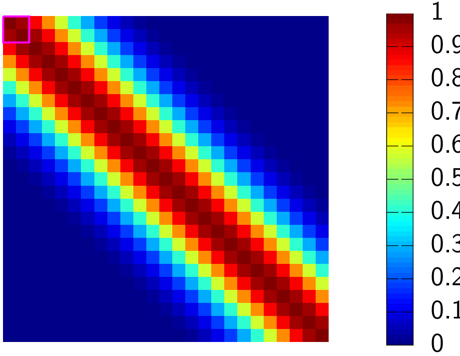
(b) colormap showing correlations between dimensions.

Figure: A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



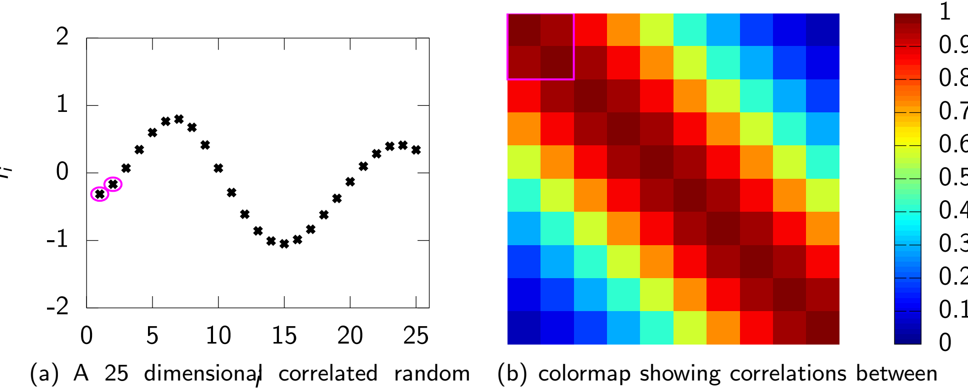
(a) A 25 dimensional correlated random variable (values ploted against index)



(b) colormap showing correlations between dimensions.

Figure: A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample

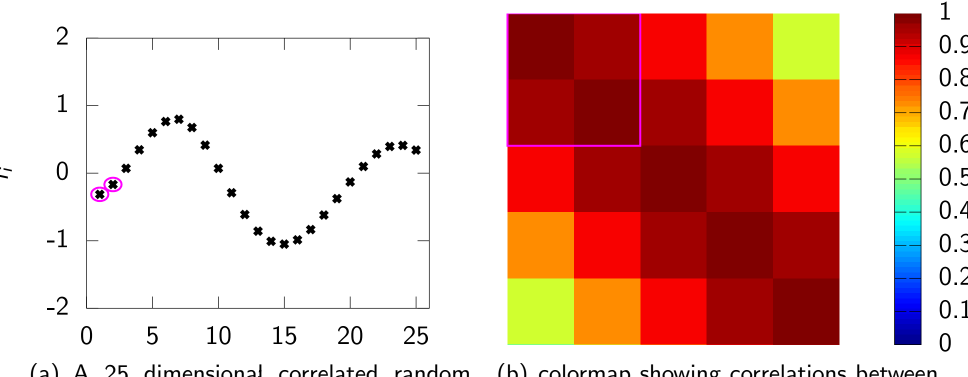


(a) A 25 dimensional correlated random variable (values ploted against index)

(b) colormap showing correlations between dimensions.

Figure: A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample

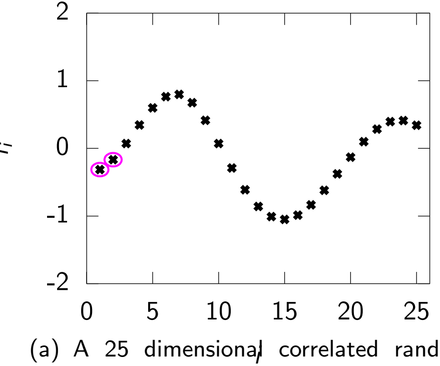


(a) A 25 dimensional correlated random variable (values ploted against index)

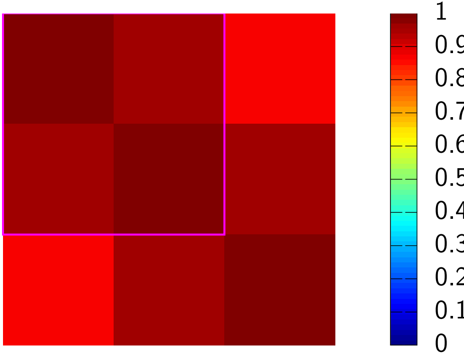
(b) colormap showing correlations between dimensions.

Figure: A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



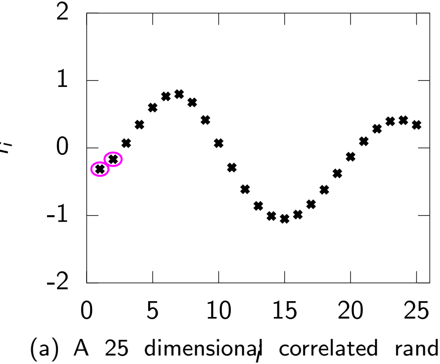
(a) A 25 dimensional correlated random variable (values ploted against index)



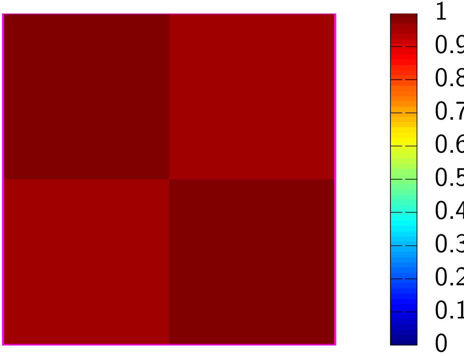
(b) colormap showing correlations between dimensions.

Figure: A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



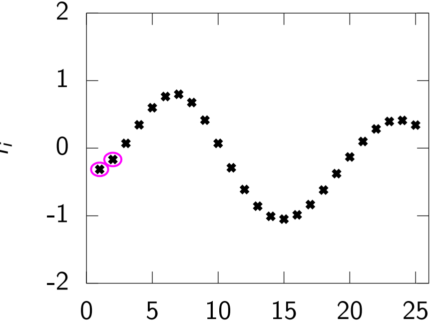
(a) A 25 dimensional correlated random variable (values ploted against index)



(b) colormap showing correlations between dimensions.

Figure: A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



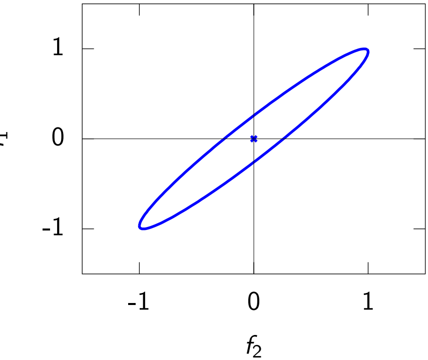
(a) A 25 dimensional correlated random variable (values ploted against index)

$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

(b) correlation between f_1 and f_2.

Figure: A sample from a 25 dimensional Gaussian distribution.

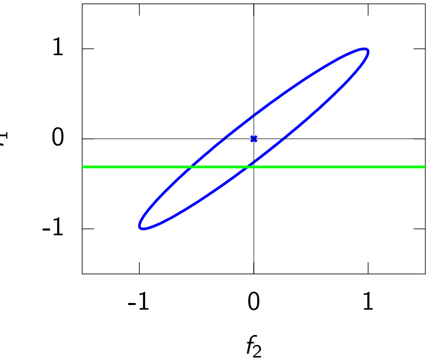
Prediction of f_2 from f_1



$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_2)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_2|f_1 = -0.313)$.

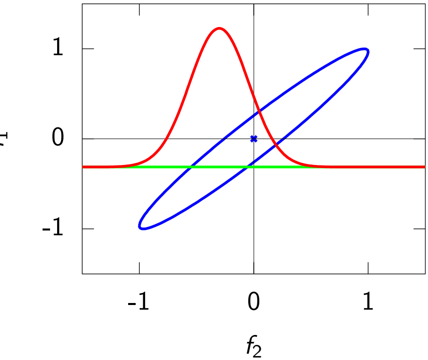
Prediction of f_2 from f_1



$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_2)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_2|f_1 = -0.313)$.

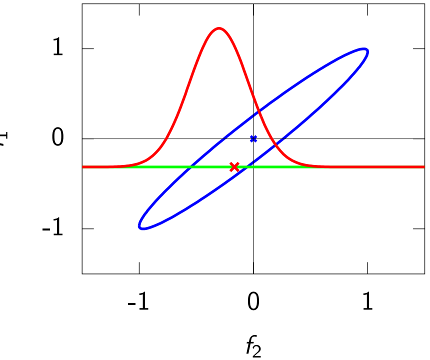
Prediction of f_2 from f_1



$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_2)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_2|f_1 = -0.313)$.

Prediction of f_2 from f_1



$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_2)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_2|f_1 = -0.313)$.

Prediction with Correlated Gaussians

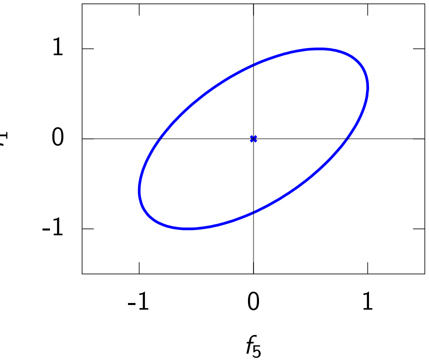
- Prediction of f_2 from f_1 requires *conditional density*.
- Conditional density is *also* Gaussian.

$$p(f_2|f_1) = \mathcal{N}\left(f_2|\frac{k_{1,2}}{k_{1,1}}f_1, k_{2,2} - \frac{k_{1,2}^2}{k_{1,1}}\right)$$

where covariance of joint density is given by

$$\mathbf{K} = \begin{bmatrix} k_{1,1} & k_{1,2} \\ k_{2,1} & k_{2,2} \end{bmatrix}$$

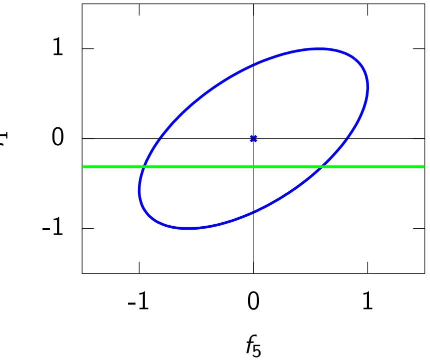
Prediction of f_5 from f_1



$$\begin{bmatrix} 1 & 0.57375 \\ 0.57375 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_5)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_5|f_1 = -0.313)$.

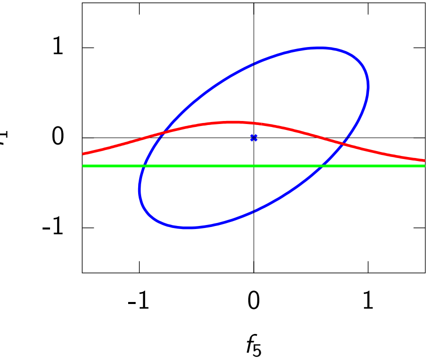
Prediction of f_5 from f_1



$$\begin{bmatrix} 1 & 0.57375 \\ 0.57375 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_5)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_5|f_1 = -0.313)$.

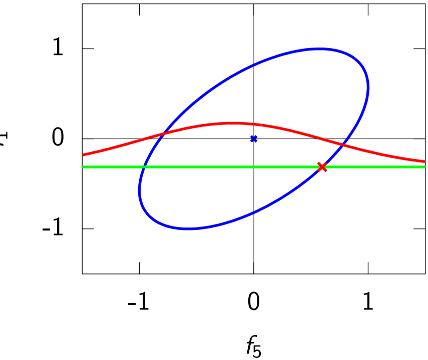
Prediction of f_5 from f_1



$$\begin{bmatrix} 1 & 0.57375 \\ 0.57375 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_5)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_5|f_1 = -0.313)$.

Prediction of f_5 from f_1



$$\begin{bmatrix} 1 & 0.57375 \\ 0.57375 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_5)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_5|f_1 = -0.313)$.

Prediction with Correlated Gaussians

- Prediction of $\mathbf{f}_*$ from $\mathbf{f}$ requires multivariate *conditional density*.
- Multivariate conditional density is *also* Gaussian.

$$p(\mathbf{f}_*|\mathbf{f}) = \mathcal{N}\left(\mathbf{f}_*|\mathbf{K}_{*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{f}, \mathbf{K}_{*,*} - \mathbf{K}_{*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{K}_{\mathbf{f},*}\right)$$

- Here covariance of joint density is given by

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{\mathbf{f},\mathbf{f}} & \mathbf{K}_{*,\mathbf{f}} \\ \mathbf{K}_{\mathbf{f},*} & \mathbf{K}_{*,*} \end{bmatrix}$$

Prediction with Correlated Gaussians

- Prediction of $\mathbf{f}_*$ from $\mathbf{f}$ requires multivariate *conditional density*.
- Multivariate conditional density is *also* Gaussian.

$$p(\mathbf{f}_*|\mathbf{f}) = \mathcal{N}\left(\mathbf{f}_*|\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$
$$\boldsymbol{\mu} = \mathbf{K}_{*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{f}$$
$$\boldsymbol{\Sigma} = \mathbf{K}_{*,*} - \mathbf{K}_{*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{K}_{\mathbf{f},*}$$

- Here covariance of joint density is given by

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{\mathbf{f},\mathbf{f}} & \mathbf{K}_{*,\mathbf{f}} \\ \mathbf{K}_{\mathbf{f},*} & \mathbf{K}_{*,*} \end{bmatrix}$$

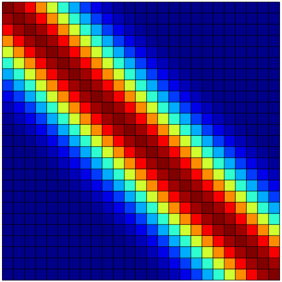
Covariance Functions

Where did this covariance matrix come from?

Exponentiated Quadratic Kernel Function (RBF, Squared Exponential, Gaussian)

$$k\left(t, t'\right) = \alpha \exp\left(-\frac{\left\|t - t'\right\|_2^2}{2\ell^2}\right)$$

- Covariance matrix is built using the *inputs* to the function t.
- For the example above it was based on Euclidean distance.
- The covariance function is also know as a kernel.



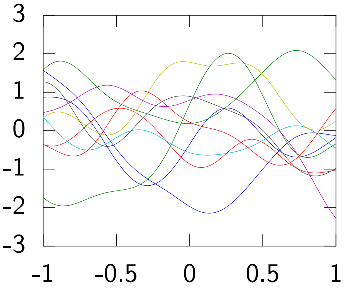
Covariance Functions

Where did this covariance matrix come from?

Exponentiated Quadratic Kernel Function (RBF, Squared Exponential, Gaussian)

$$k\left(t, t'\right) = \alpha \exp\left(-\frac{\|t - t'\|_2^2}{2\ell^2}\right)$$

- Covariance matrix is built using the *inputs* to the function t.
- For the example above it was based on Euclidean distance.
- The covariance function is also know as a kernel.



Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{\left\|t_i - t_j\right\|^2}{2\ell^2}\right)$$

$$t_1 = -3.0, \; t_1 = -3.0$$

$$k_{1,1} = 1.00 \times \exp\left(-\frac{(-3.0 - -3.0)^2}{2 \times 2.00^2}\right)$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_1 = -3.0, \; t_1 = -3.0$$

$$k_{1,1} = 1.00 \times \exp\left(-\frac{(-3.0 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & & \\ & & \\ & & \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_2 = 1.20, \; t_1 = -3.0$$

$$k_{2,1} = 1.00 \times \exp\left(-\frac{(1.20 - 1.20)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & & \\ & & \\ & & \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_2 = 1.20, \; t_1 = -3.0$$

$$k_{2,1} = 1.00 \times \exp\left(-\frac{(1.20 - 1.20)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & & \\ \boxed{0.110} & & \\ & & \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$t_2 = 1.20,\ t_1 = -3.0$

$$k_{2,1} = 1.00 \times \exp\left(-\frac{(1.20-1.20)^2}{2\times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & & \\ & & \end{bmatrix}$$

$t_1 = -3.0,\ t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_2 = 1.20, \; t_2 = 1.20$$

$$k_{2,2} = 1.00 \times \exp\left(-\frac{(1.20 - 1.20)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & & \\ & & \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_2 = 1.20, \; t_2 = 1.20$$

$$k_{2,2} = 1.00 \times \exp\left(-\frac{(1.20 - 1.20)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & \boxed{1.00} & \\ & & \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{\left\|t_i - t_j\right\|^2}{2\ell^2}\right)$$

$$t_3 = 1.40,\ t_1 = -3.0$$

$$k_{3,1} = 1.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & 1.00 & \\ & & \end{bmatrix}$$

$t_1 = -3.0,\ t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$t_3 = 1.40, \; t_1 = -3.0$

$k_{3,1} = 1.00 \times \exp\left(-\frac{(1.40-1.40)^2}{2\times 2.00^2}\right)$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & 1.00 & \\ 0.0889 & & \end{bmatrix}$$

$t_1 = -3.0, \; t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{\left\|t_i - t_j\right\|^2}{2\ell^2}\right)$$

$$t_3 = 1.40,\ t_1 = -3.0$$

$$k_{3,1} = 1.00 \times \exp\left(-\frac{(1.40-1.40)^2}{2\times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & \\ 0.0889 & & \end{bmatrix}$$

$t_1 = -3.0,\ t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_3 = 1.40, \ t_2 = 1.20$$

$$k_{3,2} = 1.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & \\ 0.0889 & & \end{bmatrix}$$

$t_1 = -3.0, \ t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_3 = 1.40, \; t_2 = 1.20$$

$$k_{3,2} = 1.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & \\ 0.0889 & \boxed{0.995} & \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_3 = 1.40,\ t_2 = 1.20$$

$$k_{3,2} = 1.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & 0.995 \\ 0.0889 & 0.995 & \end{bmatrix}$$

$t_1 = -3.0,\ t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$t_3 = 1.40, \; t_3 = 1.40$

$$k_{3,3} = 1.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & 0.995 \\ 0.0889 & 0.995 & \end{bmatrix}$$

$t_1 = -3.0, \; t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{\left\|t_i - t_j\right\|^2}{2\ell^2}\right)$$

$$t_3 = 1.40, \; t_3 = 1.40$$

$$k_{3,3} = 1.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & 0.995 \\ 0.0889 & 0.995 & \boxed{1.00} \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

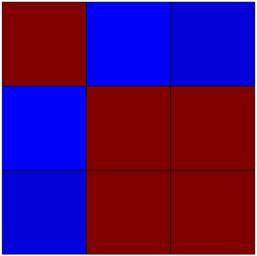
Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{\left|\left|t_i - t_j\right|\right|^2}{2\ell^2}\right)$$

$$t_3 = 1.40, \; t_3 = 1.40$$

$$k_{3,3} = 1.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 2.00^2}\right)$$



$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{\left|\left|t_i - t_j\right|\right|^2}{2\ell^2}\right)$$

$$t_1 = -3,\ t_1 = -3$$

$$k_{1,1} = 1.0 \times \exp\left(-\frac{(-3 - -3)^2}{2 \times 2.0^2}\right)$$

$t_1 = -3,\ t_2 = 1.2,\ t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_1 = -3, \; t_1 = -3$$

$$k_{1,1} = 1.0 \times \exp\left(-\frac{(-3 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} \boxed{1.0} & & & \\ & & & \\ & & & \\ & & & \end{bmatrix}$$

$t_1 = -3, \; t_2 = 1.2, \; t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_2 = 1.2,\ t_1 = -3$$

$$k_{2,1} = 1.0 \times \exp\left(-\frac{(1.2-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & & & \\ & & & \\ & & & \\ & & & \end{bmatrix}$$

$t_1 = -3,\ t_2 = 1.2,\ t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_2 = 1.2,\ t_1 = -3$$

$$k_{2,1} = 1.0 \times \exp\left(-\frac{(1.2 - 1.2)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & & & \\ 0.11 & & & \\ & & & \\ & & & \end{bmatrix}$$

$t_1 = -3$, $t_2 = 1.2$, $t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$t_2 = 1.2,\ t_1 = -3$

$k_{2,1} = 1.0 \times \exp\left(-\frac{(1.2-1.2)^2}{2\times 2.0^2}\right)$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & & & \\ & & & \\ & & & \end{bmatrix}$$

$t_1 = -3,\ t_2 = 1.2,\ t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_2 = 1.2,\ t_2 = 1.2$$

$$k_{2,2} = 1.0 \times \exp\left(-\frac{(1.2-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & & & \\ & & & \\ & & & \end{bmatrix}$$

$t_1 = -3,\ t_2 = 1.2,\ t_3 = 1.4,$ and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_2 = 1.2,\; t_2 = 1.2$$

$$k_{2,2} = 1.0 \times \exp\left(-\frac{(1.2-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & \boxed{1.0} & & \\ & & & \\ & & & \end{bmatrix}$$

$t_1 = -3,\; t_2 = 1.2,\; t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_3 = 1.4, \; t_1 = -3$$

$$k_{3,1} = 1.0 \times \exp\left(-\frac{(1.4 - 1.4)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & 1.0 & & \\ & & & \\ & & & \end{bmatrix}$$

$t_1 = -3$, $t_2 = 1.2$, $t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_3 = 1.4,\ t_1 = -3$$

$$k_{3,1} = 1.0 \times \exp\left(-\frac{(1.4 - 1.4)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & 1.0 & & \\ 0.089 & & & \\ & & & \end{bmatrix}$$

$t_1 = -3$, $t_2 = 1.2$, $t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_3 = 1.4, \; t_1 = -3$$

$$k_{3,1} = 1.0 \times \exp\left(-\frac{(1.4-1.4)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & & \\ 0.089 & & & \\ & & & \end{bmatrix}$$

$t_1 = -3$, $t_2 = 1.2$, $t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$t_3 = 1.4,\ t_2 = 1.2$

$k_{3,2} = 1.0 \times \exp\left(-\frac{(1.4-1.4)^2}{2\times 2.0^2}\right)$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & & \\ 0.089 & & & \\ & & & \end{bmatrix}$$

$t_1 = -3,\ t_2 = 1.2,\ t_3 = 1.4,$ and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$t_3 = 1.4,\ t_2 = 1.2$

$k_{3,2} = 1.0 \times \exp\left(-\frac{(1.4-1.4)^2}{2\times 2.0^2}\right)$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & & \\ 0.089 & \boxed{1.0} & & \\ & & & \end{bmatrix}$$

$t_1 = -3,\ t_2 = 1.2,\ t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$t_3 = 1.4, \; t_2 = 1.2$

$$k_{3,2} = 1.0 \times \exp\left(-\frac{(1.4-1.4)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & & \\ & & & \end{bmatrix}$$

$t_1 = -3, \; t_2 = 1.2, \; t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_3 = 1.4, \ t_3 = 1.4$$

$$k_{3,3} = 1.0 \times \exp\left(-\frac{(1.4-1.4)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & & \\ & & & \end{bmatrix}$$

$t_1 = -3, \ t_2 = 1.2, \ t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$t_3 = 1.4, \ t_3 = 1.4$

$$k_{3,3} = 1.0 \times \exp\left(-\frac{(1.4-1.4)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 \\ 0.11 & 1.0 & 1.0 \\ 0.089 & 1.0 & \boxed{1.0} \end{bmatrix}$$

$t_1 = -3, \ t_2 = 1.2, \ t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_4 = 2.0, \; t_1 = -3$$

$$k_{4,1} = 1.0 \times \exp\left(-\frac{(2.0 - 2.0)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ & & & \end{bmatrix}$$

$t_1 = -3, \; t_2 = 1.2, \; t_3 = 1.4,$ and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0.$

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_4 = 2.0,\ t_1 = -3$$

$$k_{4,1} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ \boxed{0.044} & & & \end{bmatrix}$$

$t_1 = -3$, $t_2 = 1.2$, $t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_4 = 2.0, \; t_1 = -3$$

$$k_{4,1} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & & & \end{bmatrix}$$

$t_1 = -3, \; t_2 = 1.2, \; t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_4 = 2.0,\ t_2 = 1.2$$

$$k_{4,2} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & & & \end{bmatrix}$$

$t_1 = -3,\ t_2 = 1.2,\ t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$t_4 = 2.0, \; t_2 = 1.2$

$k_{4,2} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2\times 2.0^2}\right)$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & \boxed{0.92} & & \end{bmatrix}$$

$t_1 = -3, \; t_2 = 1.2, \; t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_4 = 2.0, \ t_2 = 1.2$$

$$k_{4,2} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & 0.92 & & \end{bmatrix}$$

$t_1 = -3, \ t_2 = 1.2, \ t_3 = 1.4,$ and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$t_4 = 2.0, \; t_3 = 1.4$

$k_{4,3} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2\times 2.0^2}\right)$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & 0.92 & & \end{bmatrix}$$

$t_1 = -3, \; t_2 = 1.2, \; t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_4 = 2.0,\ t_3 = 1.4$$

$$k_{4,3} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & 0.92 & \boxed{0.96} & \end{bmatrix}$$

$t_1 = -3,\ t_2 = 1.2,\ t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_4 = 2.0, \; t_3 = 1.4$$

$$k_{4,3} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & 0.96 \\ 0.044 & 0.92 & 0.96 & \end{bmatrix}$$

$t_1 = -3, \; t_2 = 1.2, \; t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_4 = 2.0, \; t_4 = 2.0$$

$$k_{4,4} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & 0.96 \\ 0.044 & 0.92 & 0.96 & \end{bmatrix}$$

$t_1 = -3$, $t_2 = 1.2$, $t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_4 = 2.0, \; t_4 = 2.0$$

$$k_{4,4} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & 0.96 \\ 0.044 & 0.92 & 0.96 & \boxed{1.0} \end{bmatrix}$$

$t_1 = -3, \; t_2 = 1.2, \; t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

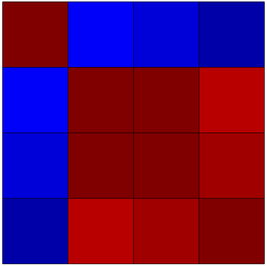
Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$t_4 = 2.0, \; t_4 = 2.0$

$$k_{4,4} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2 \times 2.0^2}\right)$$



$t_1 = -3, \; t_2 = 1.2, \; t_3 = 1.4$, and $t_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{\left|\left|t_i - t_j\right|\right|^2}{2\ell^2}\right)$$

$$t_1 = -3.0,\ t_1 = -3.0$$

$$k_{1,1} = 4.00 \times \exp\left(-\frac{(-3.0 - -3.0)^2}{2 \times 5.00^2}\right)$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_1 = -3.0, \; t_1 = -3.0$$

$$k_{1,1} = 4.00 \times \exp\left(-\frac{(-3.0 - -3.0)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} \boxed{4.00} & & \\ & & \\ & & \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_2 = 1.20,\ t_1 = -3.0$$

$$k_{2,1} = 4.00 \times \exp\left(-\frac{(1.20 - 1.20)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & & \\ & & \\ & & \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{\left|\left|t_i - t_j\right|\right|^2}{2\ell^2}\right)$$

$$t_2 = 1.20, \; t_1 = -3.0$$

$$k_{2,1} = 4.00 \times \exp\left(-\frac{(1.20 - 1.20)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & & \\ 2.81 & & \\ & & \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_2 = 1.20,\ t_1 = -3.0$$

$$k_{2,1} = 4.00 \times \exp\left(-\frac{(1.20 - 1.20)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & \\ 2.81 & & \\ & & \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_2 = 1.20, \; t_2 = 1.20$$

$$k_{2,2} = 4.00 \times \exp\left(-\frac{(1.20 - 1.20)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & \\ 2.81 & & \\ & & \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_2 = 1.20, \; t_2 = 1.20$$

$$k_{2,2} = 4.00 \times \exp\left(-\frac{(1.20 - 1.20)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 \\ 2.81 & \boxed{4.00} \end{bmatrix}$$

$t_1 = -3.0, \; t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$t_3 = 1.40,\ t_1 = -3.0$

$$k_{3,1} = 4.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & \\ 2.81 & 4.00 & \\ & & \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_3 = 1.40, \; t_1 = -3.0$$

$$k_{3,1} = 4.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & \\ 2.81 & 4.00 & \\ \boxed{2.72} & & \end{bmatrix}$$

$t_1 = -3.0, \; t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$t_3 = 1.40, \; t_1 = -3.0$

$$k_{3,1} = 4.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & \\ 2.72 & & \end{bmatrix}$$

$t_1 = -3.0, \; t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_3 = 1.40,\ t_2 = 1.20$$

$$k_{3,2} = 4.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & \\ 2.72 & & \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_3 = 1.40,\ t_2 = 1.20$$

$$k_{3,2} = 4.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & \\ 2.72 & \boxed{4.00} & \end{bmatrix}$$

$t_1 = -3.0,\ t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_3 = 1.40,\ t_2 = 1.20$$

$$k_{3,2} = 4.00 \times \exp\left(-\frac{(1.40-1.40)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & 4.00 \\ 2.72 & 4.00 & \end{bmatrix}$$

$t_1 = -3.0,\ t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_3 = 1.40, \; t_3 = 1.40$$

$$k_{3,3} = 4.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & 4.00 \\ 2.72 & 4.00 & \end{bmatrix}$$

$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(t_i, t_j) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$t_3 = 1.40, \; t_3 = 1.40$

$$k_{3,3} = 4.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & 4.00 \\ 2.72 & 4.00 & \boxed{4.00} \end{bmatrix}$$

$t_1 = -3.0, \; t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

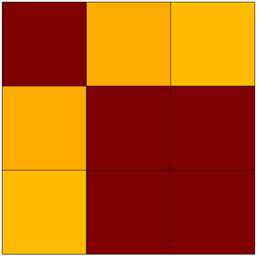
Covariance Functions

Where did this covariance matrix come from?

$$k\left(t_i, t_j\right) = \alpha \exp\left(-\frac{||t_i - t_j||^2}{2\ell^2}\right)$$

$$t_3 = 1.40, \; t_3 = 1.40$$

$$k_{3,3} = 4.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 5.00^2}\right)$$



$t_1 = -3.0$, $t_2 = 1.20$, and $t_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Gaussian Process Interpolation

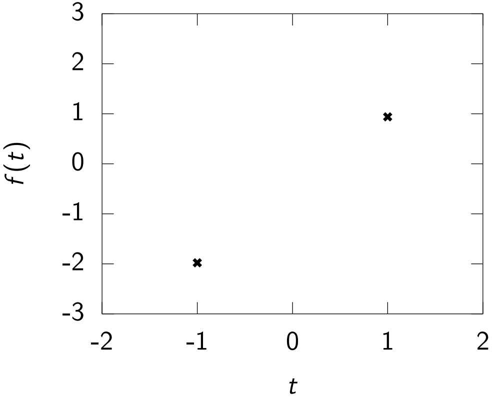


Figure: Real example: BACCO (see *e.g.* (?)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

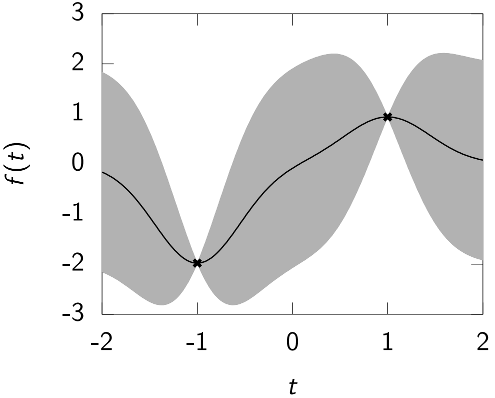


Figure: Real example: BACCO (see *e.g.* (**?**)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

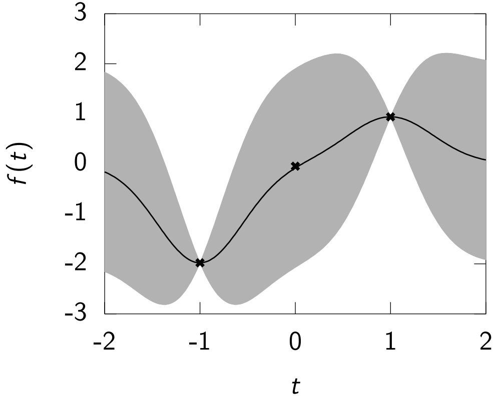


Figure: Real example: BACCO (see *e.g.* (?)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

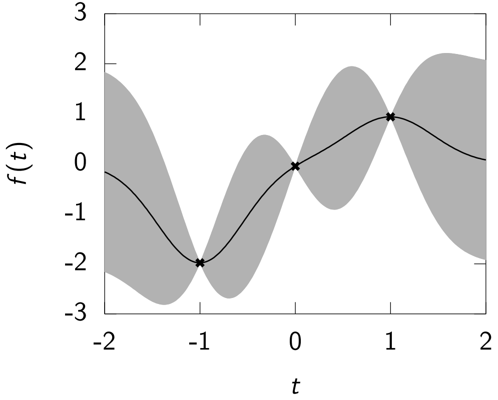


Figure: Real example: BACCO (see *e.g.* (?)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

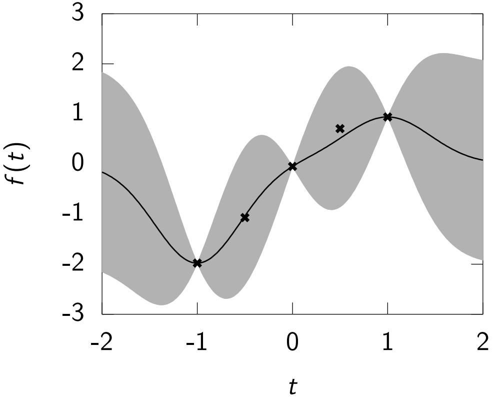


Figure: Real example: BACCO (see *e.g.* (?)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

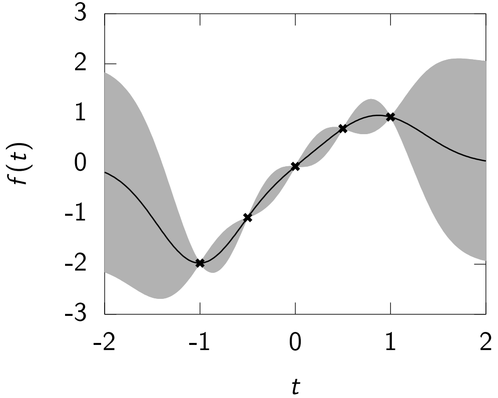


Figure: Real example: BACCO (see *e.g.* (?)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

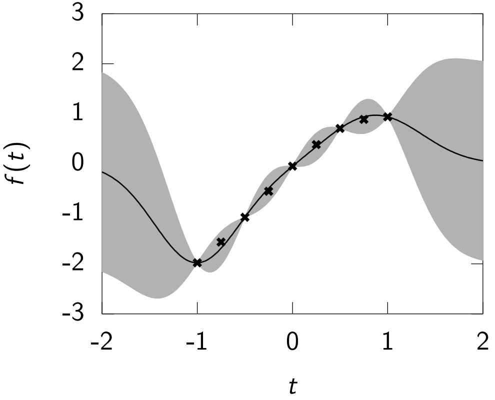


Figure: Real example: BACCO (see *e.g.* (**?**)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

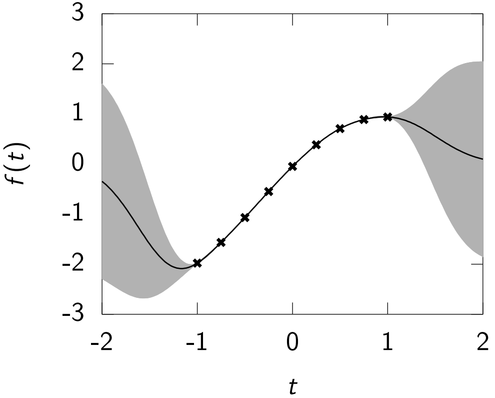


Figure: Real example: BACCO (see *e.g.* (?)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Noise Models

Graph of a GP

- Relates input variables, $\mathbf{T}$, to vector, $\mathbf{y}$, through $\mathbf{f}$ given kernel parameters $\boldsymbol{\theta}$.
- Plate notation indicates independence of $y_i|f_i$.
- Noise model, $p(y_i|f_i)$ can take several forms.
- Simplest is Gaussian noise.

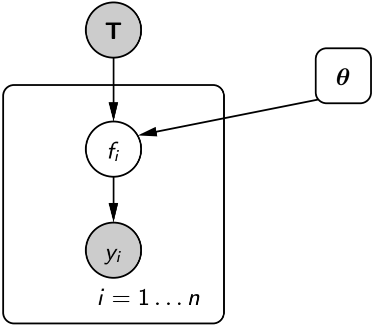


Figure: The Gaussian process depicted graphically.

Gaussian Noise

- Gaussian noise model,

$$p\left(y_i|f_i\right) = \mathcal{N}\left(y_i|f_i, \sigma^2\right)$$

where σ^2 is the variance of the noise.

- Equivalent to a covariance function of the form

$$k(t_i, t_j) = \delta_{i,j}\sigma^2$$

where $\delta_{i,j}$ is the Kronecker delta function.

- Additive nature of Gaussians means we can simply add this term to existing covariance matrices.

Gaussian Process Regression

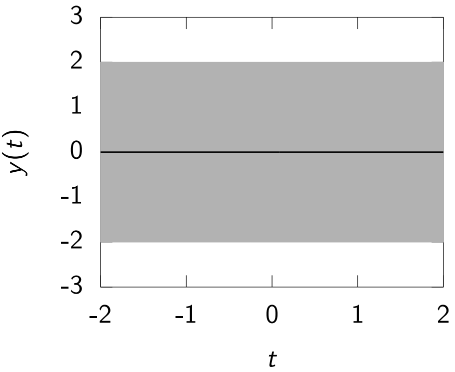


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

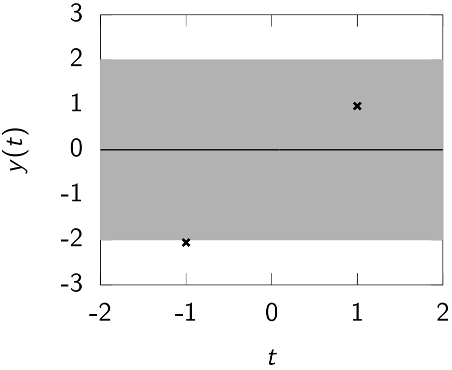


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

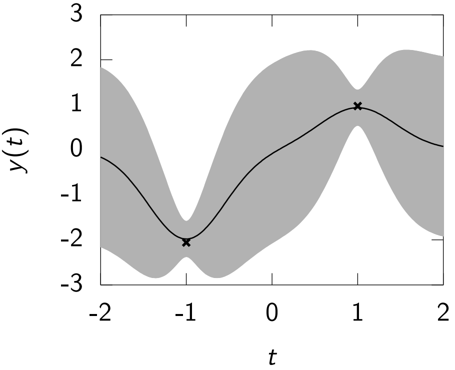


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

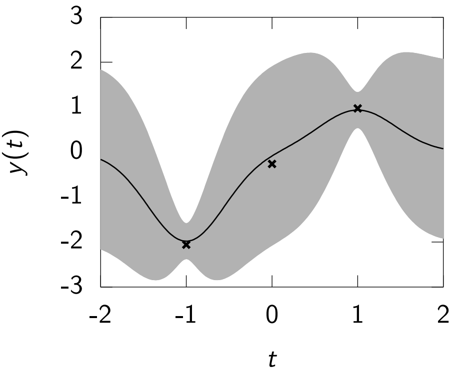


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

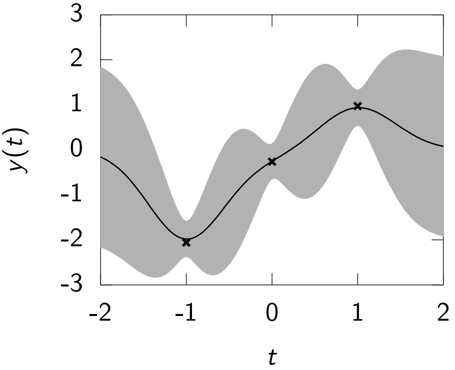


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

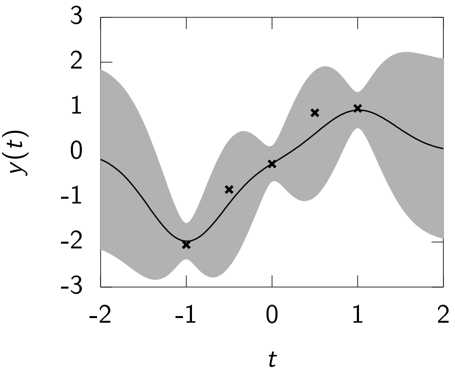


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

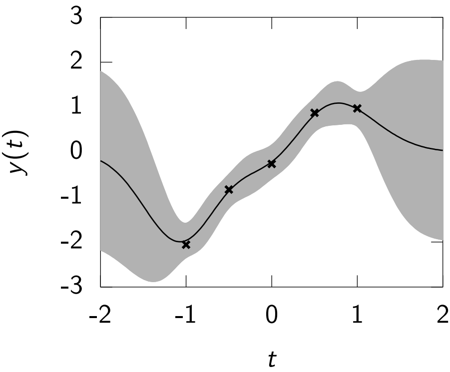


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

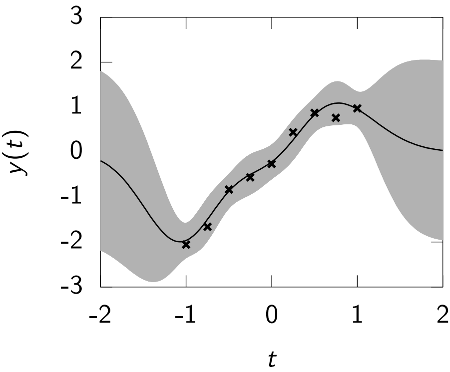


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

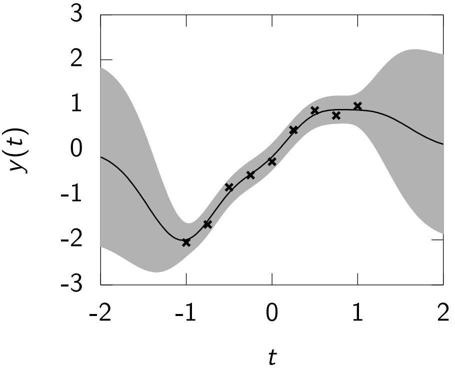


Figure: Examples include WiFi localization, C14 callibration curve.

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

$$\mathcal{N}\left(\mathbf{y}|\mathbf{0}, \mathbf{K}\right) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{K}|}\exp\left(-\frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}\right)$$

The parameters are *inside* the covariance function (matrix).

$$k_{i,j} = k(t_i, t_j; \boldsymbol{\theta})$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

$$\mathcal{N}\left(\mathbf{y}|\mathbf{0},\mathbf{K}\right) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{K}|}\exp\left(-\frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}\right)$$

The parameters are *inside* the covariance function (matrix).

$$k_{i,j} = k(t_i, t_j; \boldsymbol{\theta})$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

$$\log \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{K}| - \frac{\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y}}{2}$$

The parameters are *inside* the covariance function (matrix).

$$k_{i,j} = k(t_i, t_j; \boldsymbol{\theta})$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

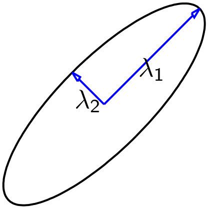
$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top\mathbf{K}^{-1}\mathbf{y}}{2}$$

The parameters are *inside* the covariance function (matrix).

$$k_{i,j} = k(t_i, t_j; \boldsymbol{\theta})$$

Eigendecomposition of Covariance

$$\mathbf{K} = \mathbf{R}\boldsymbol{\Lambda}^2\mathbf{R}^\top$$

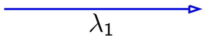


where $\boldsymbol{\Lambda}$ is a *diagonal* matrix and $\mathbf{R}^\top\mathbf{R} = \mathbf{I}$.

Useful representation since $|\mathbf{K}| = \left|\boldsymbol{\Lambda}^2\right| = |\boldsymbol{\Lambda}|^2$.

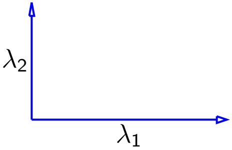
Capacity control: $\log |\mathbf{K}|$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



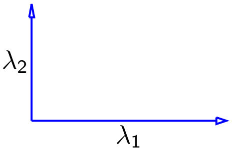
Capacity control: $\log |\mathbf{K}|$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



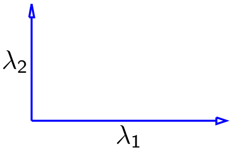
Capacity control: $\log |\mathbf{K}|$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



Capacity control: $\log |\mathbf{K}|$

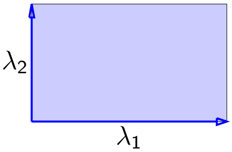
$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



$$|\mathbf{\Lambda}| = \lambda_1 \lambda_2$$

Capacity control: $\log |\mathbf{K}|$

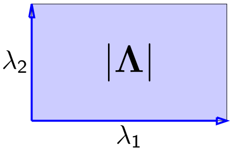
$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



$$|\boldsymbol{\Lambda}| = \lambda_1 \lambda_2$$

Capacity control: $\log |\mathbf{K}|$

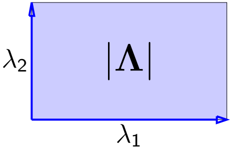
$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



$$|\mathbf{\Lambda}| = \lambda_1 \lambda_2$$

Capacity control: $\log |\mathbf{K}|$

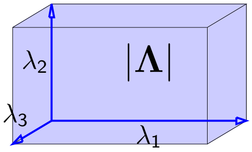
$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$



$$|\boldsymbol{\Lambda}| = \lambda_1 \lambda_2$$

Capacity control: $\log |\mathbf{K}|$

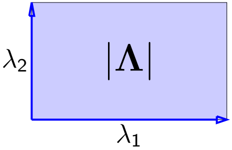
$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$



$$|\boldsymbol{\Lambda}| = \lambda_1 \lambda_2 \lambda_3$$

Capacity control: $\log |\mathbf{K}|$

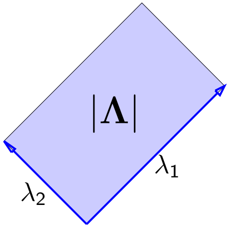
$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



$$|\mathbf{\Lambda}| = \lambda_1 \lambda_2$$

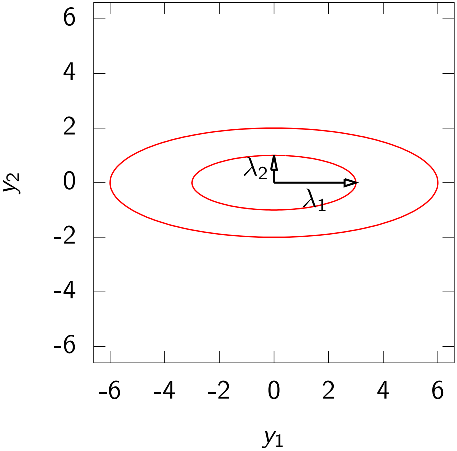
Capacity control: $\log |\mathbf{K}|$

$$\mathbf{R\Lambda} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{bmatrix}$$

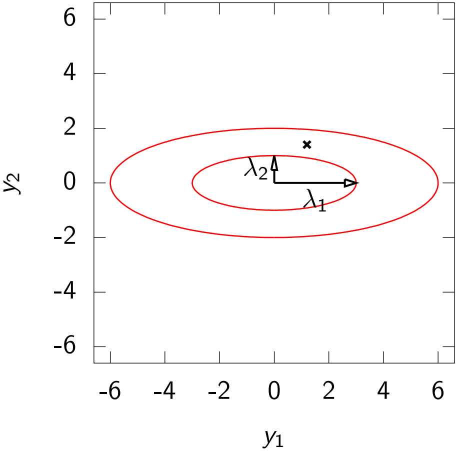


$$|\mathbf{R\Lambda}| = \lambda_1 \lambda_2$$

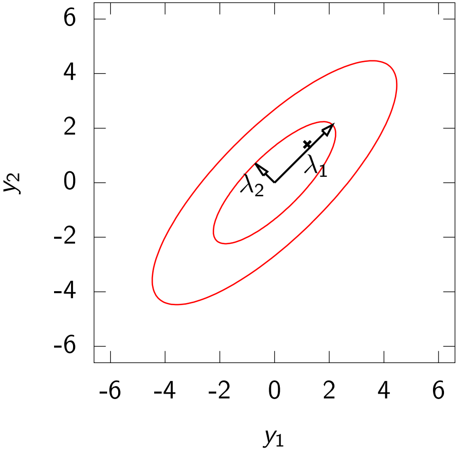
Data Fit: $\frac{\mathbf{y}^{-1}\mathbf{K}^{-1}\mathbf{y}}{2}$



Data Fit: $\frac{\mathbf{y}^{-1}\mathbf{K}^{-1}\mathbf{y}}{2}$

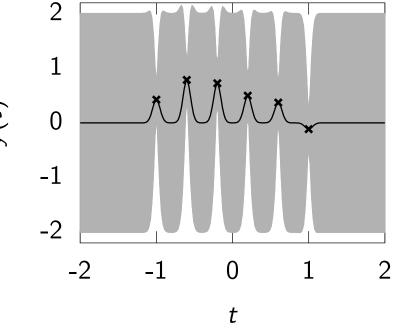


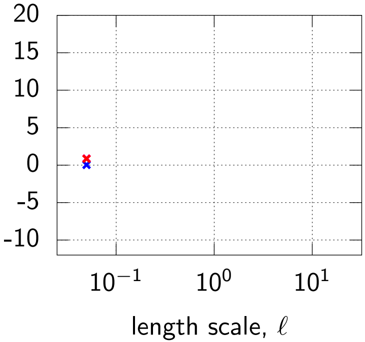
Data Fit: $\frac{\mathbf{y}^{-1}\mathbf{K}^{-1}\mathbf{y}}{2}$



Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

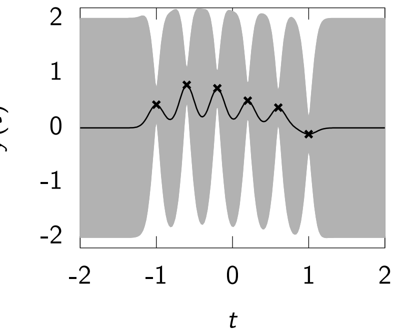


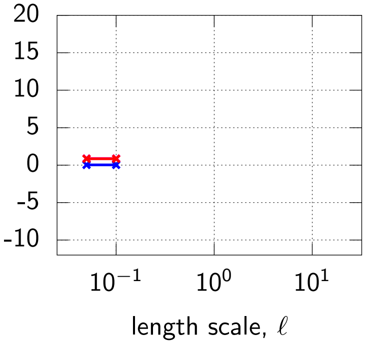


$$E(\boldsymbol{\theta}) = \frac{1}{2}\left|\mathbf{K}\right| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

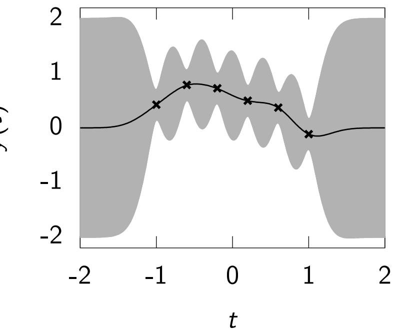


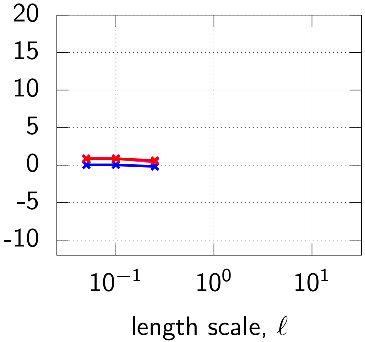


$$E(\boldsymbol{\theta}) = \frac{1}{2}\left|\mathbf{K}\right| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

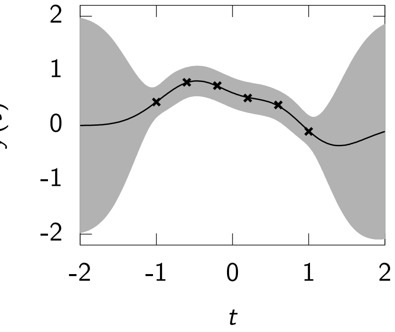


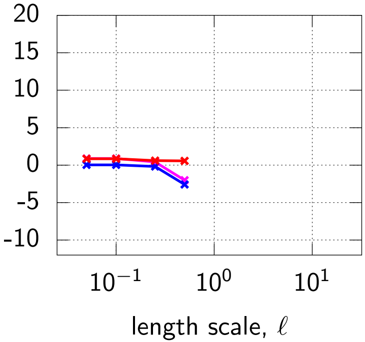


$$E(\boldsymbol{\theta}) = \frac{1}{2}\left|\mathbf{K}\right| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

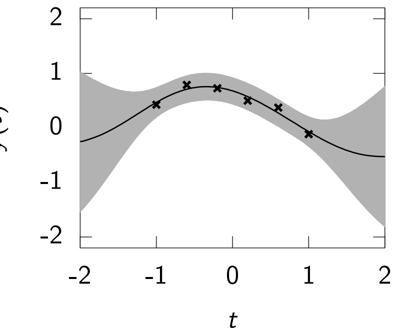


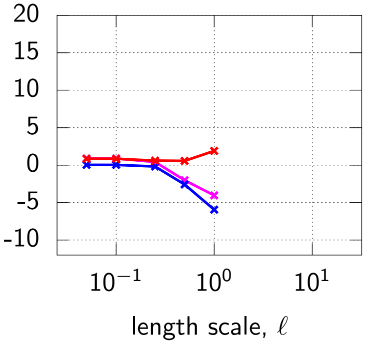


$$E(\boldsymbol{\theta}) = \frac{1}{2}\left|\mathbf{K}\right| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

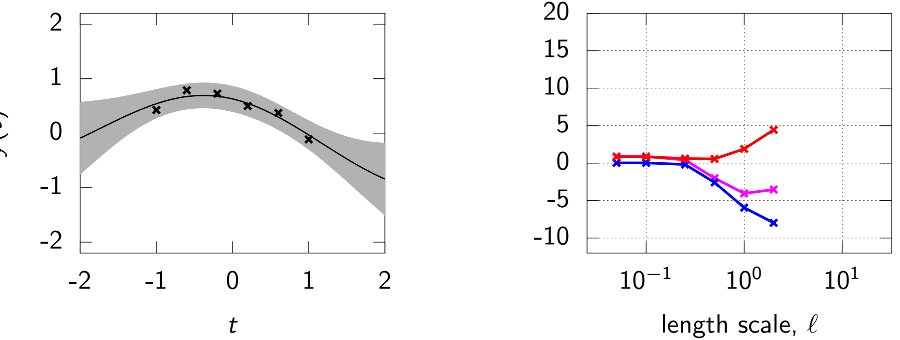




$$E(\boldsymbol{\theta}) = \frac{1}{2}\left|\mathbf{K}\right| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

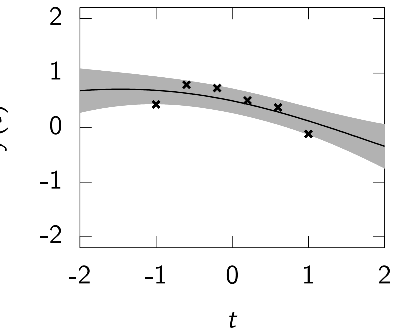
Can we determine length scales and noise levels from the data?

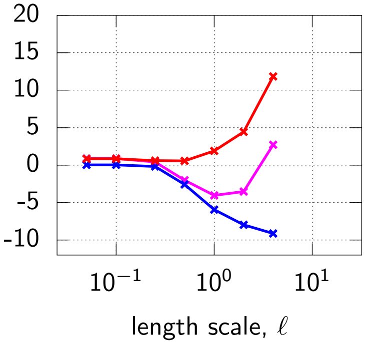


$$E(\boldsymbol{\theta}) = \frac{1}{2}\left|\mathbf{K}\right| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

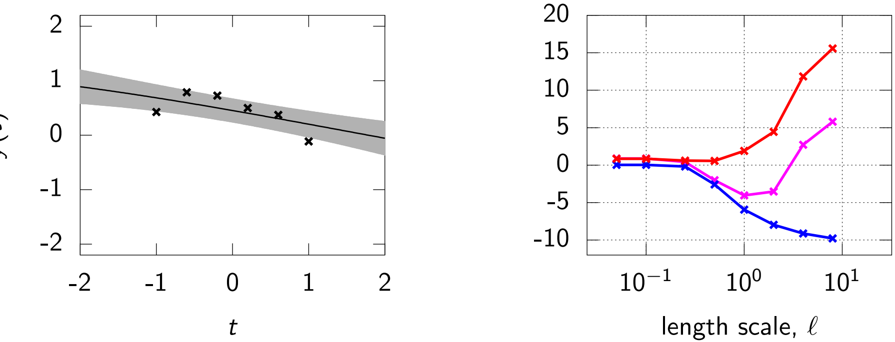




$$E(\boldsymbol{\theta}) = \frac{1}{2}\left|\mathbf{K}\right| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

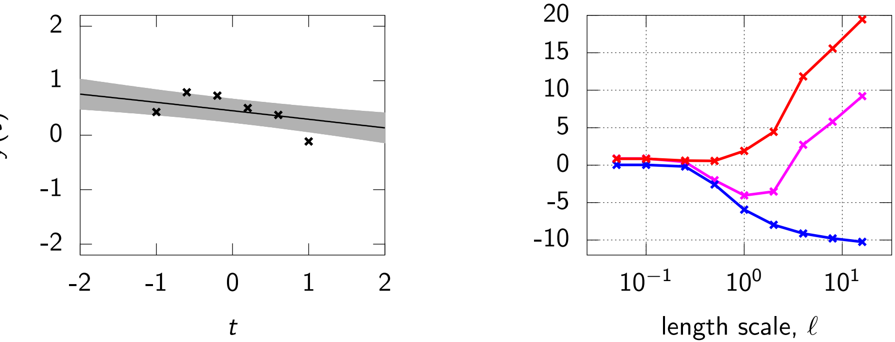
Can we determine length scales and noise levels from the data?



$$E(\boldsymbol{\theta}) = \frac{1}{2}\left|\mathbf{K}\right| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

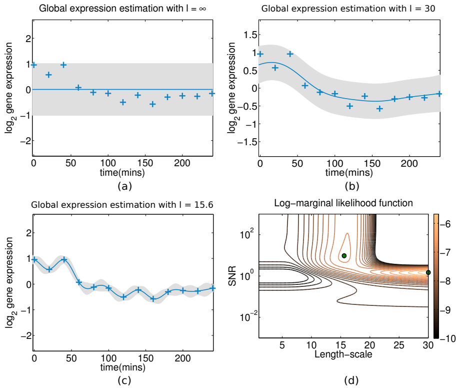
Learning Covariance Parameters

Can we determine length scales and noise levels from the data?



$$E(\boldsymbol{\theta}) = \frac{1}{2}\left|\mathbf{K}\right| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

Gene Expression Example



Data from **?**. Figure from **?**.

Outline

Limitations of Gaussian Processes

- Inference is $O(n^3)$ due to matrix inverse (in practice use Cholesky).
- Gaussian processes don't deal well with discontinuities (financial crises, phosphorylation, collisions, edges in images).
- Widely used exponentiated quadratic covariance (RBF) can be too smooth in practice (but there are many alternatives!!).

Outline

Gene Expression Example

Kalaitzis and Lawrence *BMC Bioinformatics* 2011, **12**:180
http://www.biomedcentral.com/1471-2105/12/180

BMC Bioinformatics

RESEARCH ARTICLE Open Access

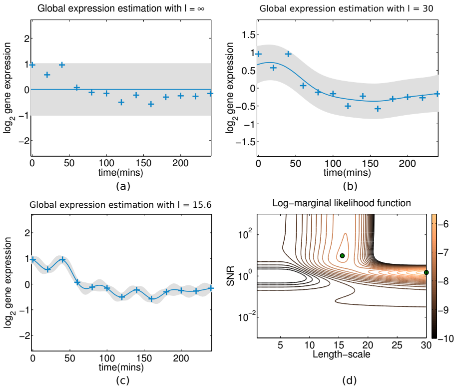
A Simple Approach to Ranking Differentially Expressed Gene Expression Time Courses through Gaussian Process Regression

Alfredo A Kalaitzis* and Neil D Lawrence*

Gene Expression Example

- Detect ‘quiet genes’ in time series.
- `http://www.bioconductor.org/packages/release/bioc/html/gprege.html` (Alfredo Kalaitzis is the maintainer).

Gene Expression Example



Data from **?**. Figure from **?**.

Summary

- Flexible method for probability densities over functions.
- Covariance function is key: defines how different data interrelate.
- Problems occur if there are discontinuities in the function.

References I

G. Della Gatta, M. Bansal, A. Ambesi-Impiombato, D. Antonini, C. Missero, and D. di Bernardo. Direct targets of the trp63 transcription factor revealed by a combination of gene expression profiling and reverse engineering. *Genome Research*, 18(6): 939–948, Jun 2008. [URL]. [DOI].

A. A. Kalaitzis and N. D. Lawrence. A simple approach to ranking differentially expressed gene expression time courses through Gaussian process regression. *BMC Bioinformatics*, 12(180), 2011. [DOI].

J. Oakley and A. O'Hagan. Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, 89(4): 769–784, 2002.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, 2006. [Google Books] .