Multivariate Linear Regression

MLAI Lecture 5

Neil D. Lawrence

Department of Computer Science
Sheffield University

5th October 2012

Outline

Multivariate Linear Regression

Multi-dimensional Inputs

- Multivariate functions involve more than one input.
- Height might be a function of weight and gender.
- There could be other contributory factors.
- Place these factors in a feature vector $\mathbf{x}_i$.
- Linear function is now defined as

$$y(\mathbf{x}_i) = \sum_{j=1}^{q} w_j x_{i,j} + c$$

Vector Notation

- Write in vector notation,

$$y(\mathbf{x}_i) = \mathbf{w}^\top \mathbf{x}_i + c$$

- Can absorb c into $\mathbf{w}$ by assuming extra input x_0 which is always 1.

$$y(\mathbf{x}_i) = \mathbf{w}^\top \mathbf{x}_i$$

Log Likelihood for Multivariate Regression

- The likelihood of a single data point is

$$p\left(t_i|x_i\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(t_i - \mathbf{w}^\top \mathbf{x}_i\right)^2}{2\sigma^2}\right).$$

- Leading to a log likelihood for the data set of

$$L(\mathbf{w}, \sigma^2) = -\frac{N}{2}\log\sigma^2 - \frac{N}{2}\log 2\pi - \frac{\sum_{i=1}^{N}\left(t_i - \mathbf{w}^\top \mathbf{x}_i\right)^2}{2\sigma^2}.$$

- And a corresponding error function of

$$E(\mathbf{w}, \sigma^2) = \frac{N}{2}\log\sigma^2 + \frac{\sum_{i=1}^{N}\left(t_i - \mathbf{w}^\top \mathbf{x}_i\right)^2}{2\sigma^2}.$$

Expand the Brackets

$$
\begin{aligned}
E(\mathbf{w}, \sigma^2) =& \frac{N}{2} \log \sigma^2 + \frac{1}{2\sigma^2} \sum_{i=1}^{N} t_i^2 - \frac{1}{\sigma^2} \sum_{i=1}^{N} t_i \mathbf{w}^\top \mathbf{x}_i \\
& + \frac{1}{2\sigma^2} \sum_{i=1}^{N} \mathbf{w}^\top \mathbf{x}_i \mathbf{x}_i^\top \mathbf{w} + \text{const.} \\
=& \frac{N}{2} \log \sigma^2 + \frac{1}{2\sigma^2} \sum_{i=1}^{N} t_i^2 - \frac{1}{\sigma^2} \mathbf{w}^\top \sum_{i=1}^{N} \mathbf{x}_i t_i \\
& + \frac{1}{2\sigma^2} \mathbf{w}^\top \left[\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^\top \right] \mathbf{w} + \text{const.}
\end{aligned}
$$

Multivariate Derivatives

- We will need some multivariate calculus.
- For now some simple multivariate differentiation:

$$\frac{\mathrm{d}\mathbf{a}^\top\mathbf{w}}{\mathrm{d}\mathbf{w}} = \mathbf{a}$$

and

$$\frac{\mathrm{d}\mathbf{w}^\top\mathbf{A}\mathbf{w}}{\mathrm{d}\mathbf{w}} = \left(\mathbf{A} + \mathbf{A}^\top\right)\mathbf{w}$$

or if $\mathbf{A}$ is symmetric ($i.e.$ $\mathbf{A} = \mathbf{A}^\top$)

$$\frac{\mathrm{d}\mathbf{w}^\top\mathbf{A}\mathbf{w}}{\mathrm{d}\mathbf{w}} = 2\mathbf{A}\mathbf{w}.$$

Differentiate

Differentiating with respect to the vector **w** we obtain

$$\frac{\partial L(\mathbf{w}, \beta)}{\partial \mathbf{w}} = \beta \sum_{i=1}^{N} \mathbf{x}_i t_i - \beta \left[\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^\top \right] \mathbf{w}$$

Leading to

$$\mathbf{w}^* = \left[\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^\top \right]^{-1} \sum_{i=1}^{N} \mathbf{x}_i t_i,$$

Rewrite in matrix notation:

$$\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^\top = \mathbf{X}^\top \mathbf{X}$$

$$\sum_{i=1}^{N} \mathbf{x}_i t_i = \mathbf{X}^\top \mathbf{t}$$

Update Equations

- Update for $\mathbf{w}^*$.

$$\mathbf{w}^* = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{t}$$

- The equation for ${\sigma^2}^*$ may also be found

$$ {\sigma^2}^* = \frac{\sum_{i=1}^{N} \left(t_i - {\mathbf{w}^*}^\top \mathbf{x}_i\right)^2}{N}.$$

Reading

- Section 1.3 of Rogers and Girolami.

References I

S. Rogers and M. Girolami. *A First Course in Machine Learning*. CRC Press, 2011. [Google Books] .