Machine Learning and the Life Sciences: from Modelling to Medicine

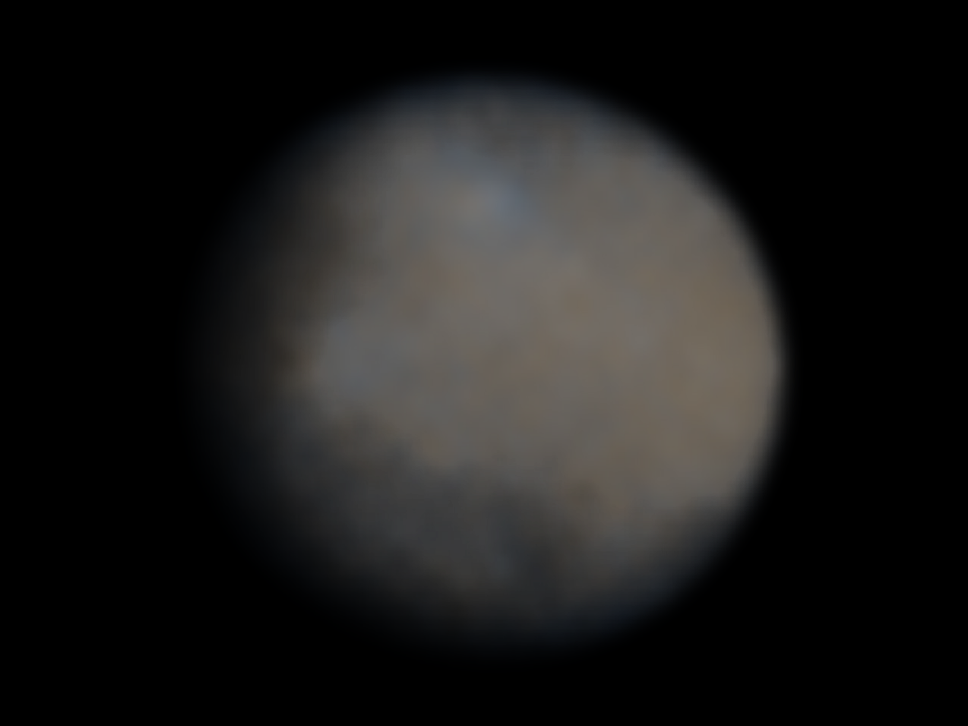
Neil D. Lawrence

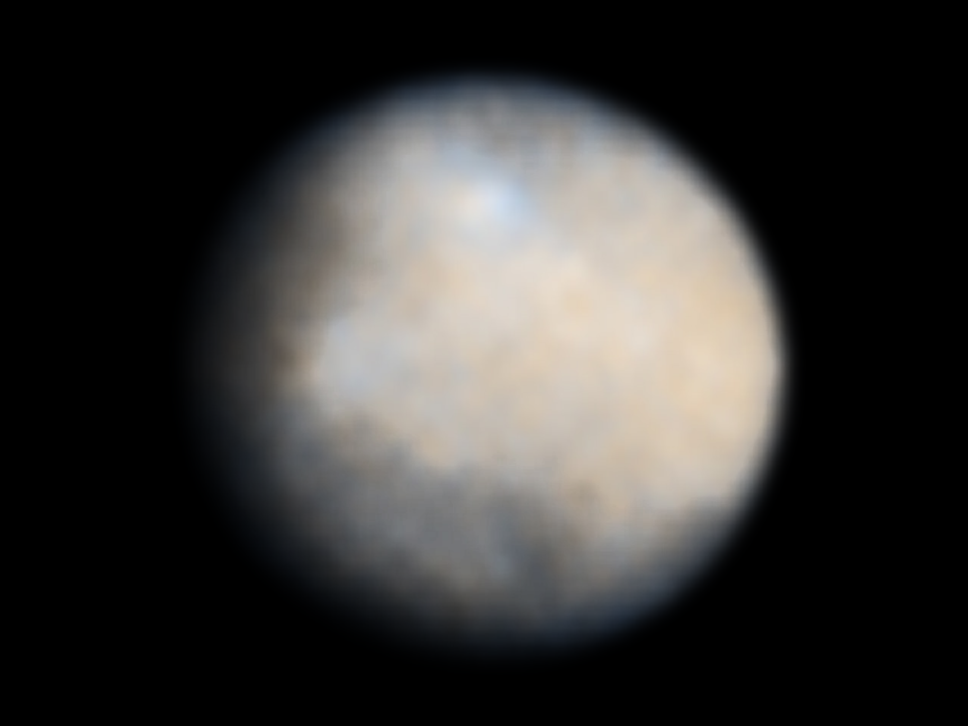
Department of Computer Science
Sheffield University

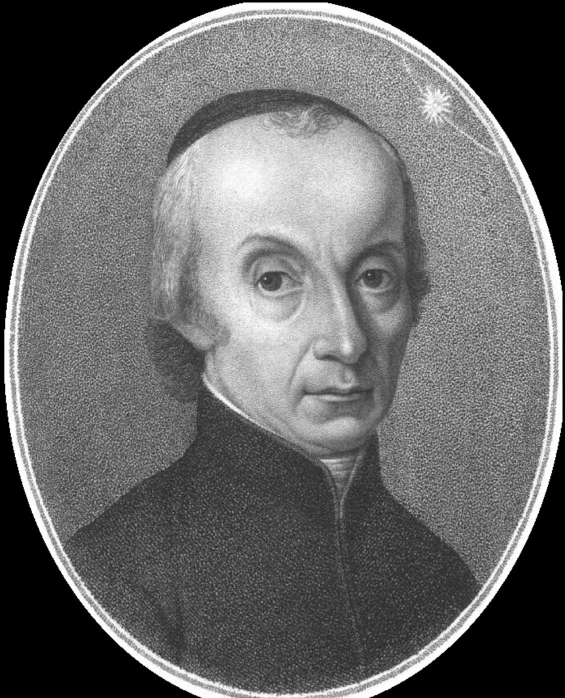
11th January 2013

Outline

What is ML?







Beobachtungen des zu Palermo d. 1. Jan. 1801 von Prof. *Piazzi* neu entdeckten Gestirns.

1801		Mittlere Sonnen-Zeit	Gerade Aufsteig in Zeit	Gerade Aufsteigung in Graden	Nördl. Abweich.	Geocentrische Länge	Geocentr. Breite	Ort der Sonne + 20" Aberration	Logar. d. Distanz ⊙ ♂
		St ' "	St ' "	° ' "	° ' "	Z ° ' "	° ' "	Z ° ' "	
Jan.	1	8 43 17,8	3 27 11,25	51 47 48,8	15 37 43,5	1 23 22 58,3	3 6 42,1	9 11 1 30,9	9,9926156
	2	8 39 4,6	3 26 53,85	51 43 27,8	15 41 5,5	1 23 19 44,3	3 2 24,9	9 12 2 28,6	9,9926317
	3	8 34 53,3	3 26 38,4::	51 39 36,0	15 44 31,6	1 23 16 58,6	2 58 9,9	9 13 3 26,6	9,9926324
	4	8 30 42,1	3 26 23,16	51 35 47,3	15 47 57,6	1 23 14 15,5	2 53 55,6	9 14 4 24,9	9,9926418
	10	8 6 15,8	3 25 32,1::	51 23 1,5	16 10 32,0	1 23 7 59,1	2 29 0,6	9 20 10 17,5	9,9927641
	11	8 2 17,5	3 25 29,73	51 22 26,0	. . . .	. . . .	. . . . . .	. . . .	. . . . .
	13	7 54 26,2	3 25 30,30	51 22 34,5	16 22 49,5	1 23 10 27,6	2 16 59,7	9 23 12 13,8	9,9928490
	14	7 50 31,7	3 25 31,72	51 22 55,8	16 27 5,7	1 23 12 1,2	2 12 56,7	9 24 14 13,5	9,9928809
	17	. . . . . .	. . . .	. . . .	16 40 13,0	. . . .	. . . . . .	. . . .	. . . . .
	18	7 35 11,3	3 25 55,::	51 28 45,0	. . . .	. . . .	. . . . . .	. . . .	. . . . .
	19	7 31 28,5	3 26 8,15	51 32 2,3	16 49 16,1	1 23 25 59,2	1 53 38,2	9 29 19 53,8	9,9930607
	21	7 24 2,7	3 26 34,27	51 38 34,1	16 58 35,9	1 23 34 21,3	1 46 6,0	10 1 20 40,3	9,9931434
	22	7 20 21,7	3 26 49,42	51 42 21,3	17 3 18,5	1 23 39 1,8	1 42 28,1	10 2 21 32,0	9,9931886
	23	7 16 43,5	3 27 6,90	51 46 43,5	17 8 5,5	1 23 44 15,7	1 38 52,1	10 3 22 22,7	9,9932348
	28	6 58 51,3	3 28 54,55	52 13 38,3	17 32 54,1	1 24 15 15,7	1 21 6,9	10 8 26 20,1	9,9935062
	30	6 51 52,9	3 29 48,14	52 27 2,1	17 43 11,0	1 24 30 9,0	1 14 16,0	10 10 27 46,2	9,9936332
	31	6 48 26,4	3 30 17,25	52 34 18,8	17 48 21,5	1 24 38 7,3	1 10 54,6	10 11 28 28,5	9,9937007
Febr.	1	6 44 59,9	3 30 47,2::	52 41 48,0	17 53 36,3	1 24 46 19,3	1 7 30,9	10 12 29 9,6	9,9937703
	2	6 41 35,8	3 31 19,06	52 49 45,9	17 58 57,5	1 24 54 57,9	1 4 1[illegible],5	10 13 29 49,9	9,9938423
	5	6 31 31,5	3 33 2,70	53 15 40,5	18 15 1,0	1 25 22 43,4	0 54 23,9	10 16 31 45,5	9,9940751
	8	6 21 39,2	3 34 58,50	53 44 37,5	18 31 23,2	1 25 53 29,5	0 45 5,0	10 19 33 33,3	9,9943276
	11	6 11 58,2	3 37 6,54	54 16 38,1	18 47 58,8	1 26 26 40,0	0 36 2,9	10 22 35 11,4	9,9945823


1801/01/01
1801/01/04
1801/01/10
1801/01/13
1801/01/19
1801/01/22
1801/01/28
1801/01/31
1801/02/05
1801/02/08
1801/02/11

hier in der Nähe der Quadratur der Einfluſs der Sonnen-Länge geringer iſt, als in andern Lagen. Dr. *Gauſs* glaubt daher, daſs es nicht undienlich wäre, wenn man die Fehler der Sonnentafeln aus ſehr genauen Beobachtungen für dieſe Zeiten beſtimmte, und die Örter der Sonne hiernach verbeſſerte. Dieſe *vierten* Elemente ſind nun folgende:

		Hieraus:	
Sonnenferne . .	326° 27′ 38″	gröſste Mittelp. Gleichung	9° 27′ 41″
☊	81 0 44	tägl. mittlere helioc. tropiſche Beweg.	770,″914
Neigung . .	10 36 57		
Log. halb. gr. Axe	0,4420527		
Excentricität . .	0,0825017		
Epoche 1800 31 Dec.	77° 36′ 34″		

Aus dieſen Elementen hat Dr. *Gauſs* folgende Örter der *Ceres Ferdinandea* im voraus berechnet. Die Zeit iſt mittlere für Mitternacht in *Palermo*.

1801	Geocentriſche Länge	Geocentriſche Breite nördl.	Logarith. des Abſtandes von der ♁	Logarith. des Abſtandes von der ☉	Verhältniſs der geſehenen Helligk.
	Z ° ′	° ′			
Nov. 25	5 20 16	9 25	0, 42181	0, 40468	0, 6102
Dec. 1	5 22 15	9 48	0, 40940	0, 40472	0, 6459
7	5 24 7	10 12	0, 39643	0, 40479	0, 6855
13	5 25 51	10 37	0, 38296	0, 40488	0, 7290
19	5 27 27	11 4	0, 36902	0, 40499	0, 7770
25	5 28 53	11 32	0, 35468	0, 40512	0, 8295
31	6 0 10	12 1	0, 34000	0, 40528	0, 8869

Sollte man den Ort des Planeten nach dieſen Elementen genauer, oder auf eine längere Zeit berechnen wollen: ſo ſetzen wir zu dieſem Behufe noch folgende Formeln hierher:

1) Zur

Epoche 1800 31 Dec. 77° 36′ 34″

Aus diesen Elementen hat Dr. *Gauſs* folgende Örter der *Ceres Ferdinandea* im voraus berechnet. Die Zeit iſt mittlere für Mitternacht in *Palermo*.

1801		Geocentriſche Länge			Geocentriſche Breite nördl.		Logarith. des Abſtandes von der ♁	Logarith. des Abſtandes von der ☉	Verhältniſs der geſehenen Helligk.
		Z	°	′	°	′			
Nov.	25	5	20	16	9	25	0, 42181	0, 40468	0, 6102
Dec.	1	5	22	15	9	48	0, 40940	0, 40472	0, 6459
	7	5	24	7	10	12	0, 39643	0, 40479	0, 6855
	13	5	25	51	10	37	0, 38296	0, 40488	0, 7290
	19	5	27	27	11	4	0, 36902	0, 40499	0, 7770
	25	5	28	53	11	32	0, 35468	0, 40512	0, 8295
	31	6	0	10	12	1	0, 34000	0, 40528	0, 8869

Sollte man den Ort des Planeten nach dieſen Elementen genauer, oder auf eine längere Zeit berechnen wollen: ſo ſetzen wir zu dieſem Behufe noch folgende Formeln hierher:

DELLA SCOPERTA

DEL NUOVO PIANETA

CERERE FERDINANDEA

OTTAVO TRA I PRIMARJ DEL NOSTRO SISTEMA SOLARE.

PALERMO
1802

NELLA STAMPERIA REALE.

What is Machine Learning?

data

- data: observations, could be actively or passively acquired (meta-data).
- model: assumptions, based on previous experience (other data! transfer learning etc), or beliefs about the regularities of the universe. Inductive bias.
- prediction: an action to be taken or a categorization or a quality score.

What is Machine Learning?

data $+$

- data: observations, could be actively or passively acquired (meta-data).
- model: assumptions, based on previous experience (other data! transfer learning etc), or beliefs about the regularities of the universe. Inductive bias.
- prediction: an action to be taken or a categorization or a quality score.

What is Machine Learning?

data $+$ model

- data: observations, could be actively or passively acquired (meta-data).
- model: assumptions, based on previous experience (other data! transfer learning etc), or beliefs about the regularities of the universe. Inductive bias.
- prediction: an action to be taken or a categorization or a quality score.

What is Machine Learning?

data + model =

- data: observations, could be actively or passively acquired (meta-data).
- model: assumptions, based on previous experience (other data! transfer learning etc), or beliefs about the regularities of the universe. Inductive bias.
- prediction: an action to be taken or a categorization or a quality score.

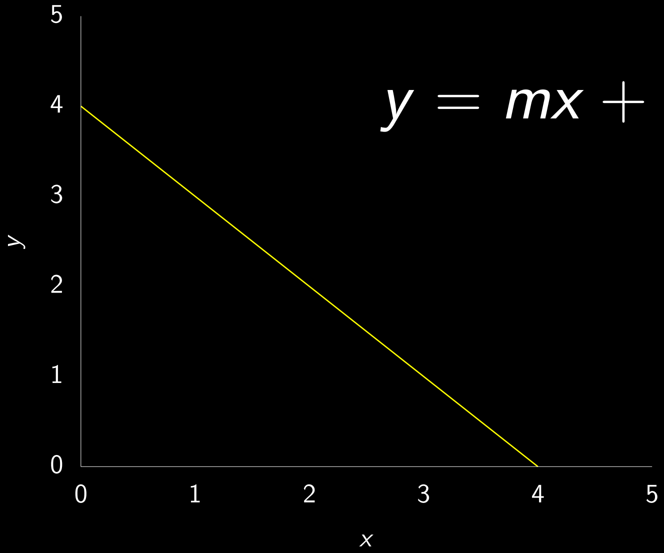
What is Machine Learning?

data + model = prediction

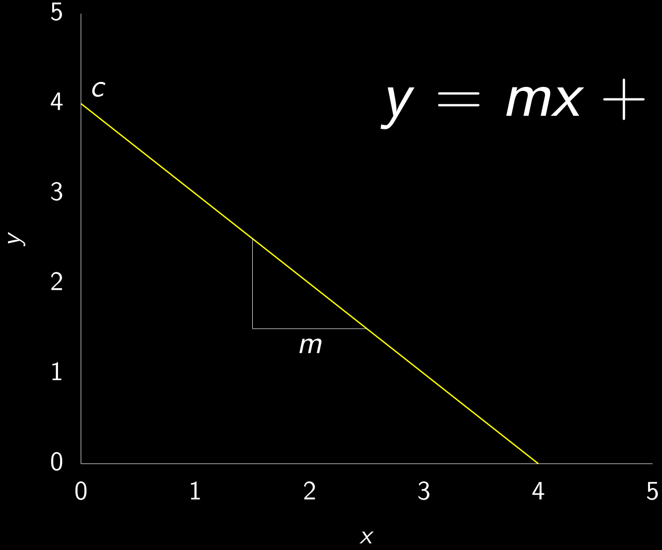
- data: observations, could be actively or passively acquired (meta-data).
- model: assumptions, based on previous experience (other data! transfer learning etc), or beliefs about the regularities of the universe. Inductive bias.
- prediction: an action to be taken or a categorization or a quality score.

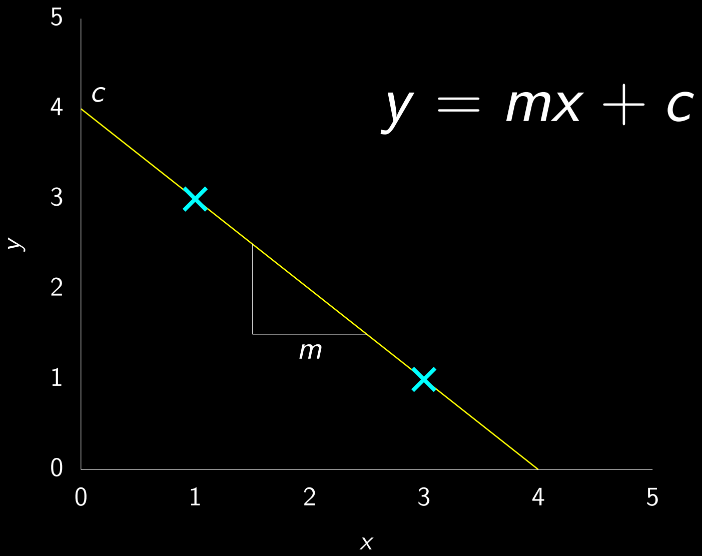
$$y = mx + c$$

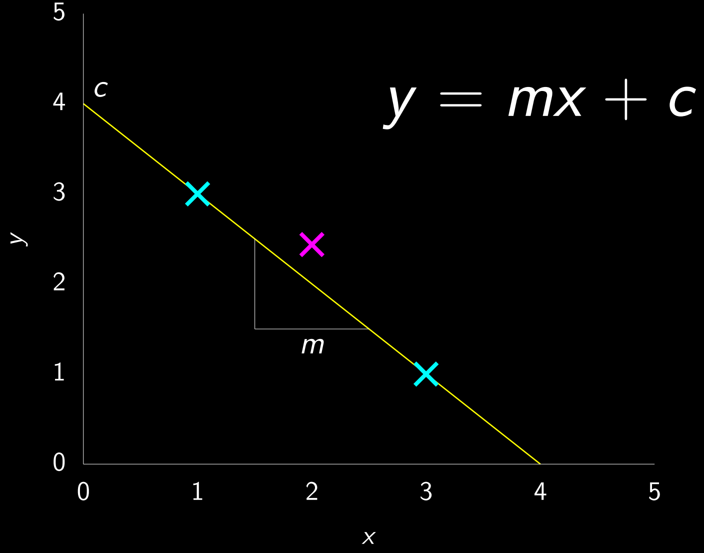
$$y = mx + c$$

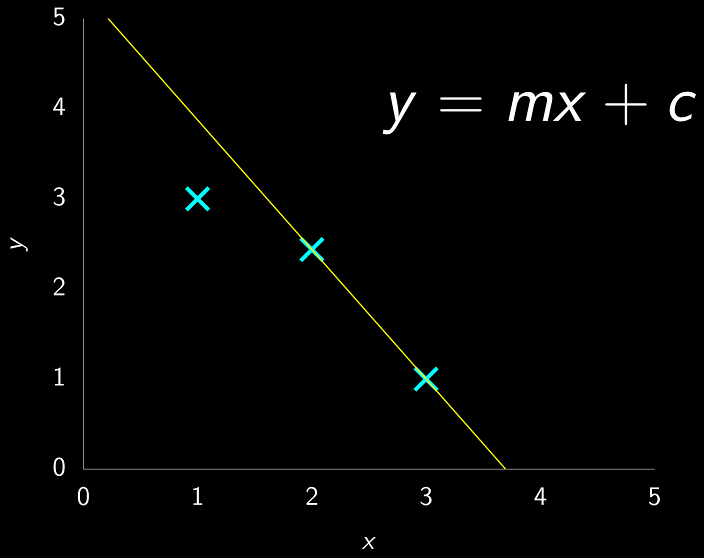


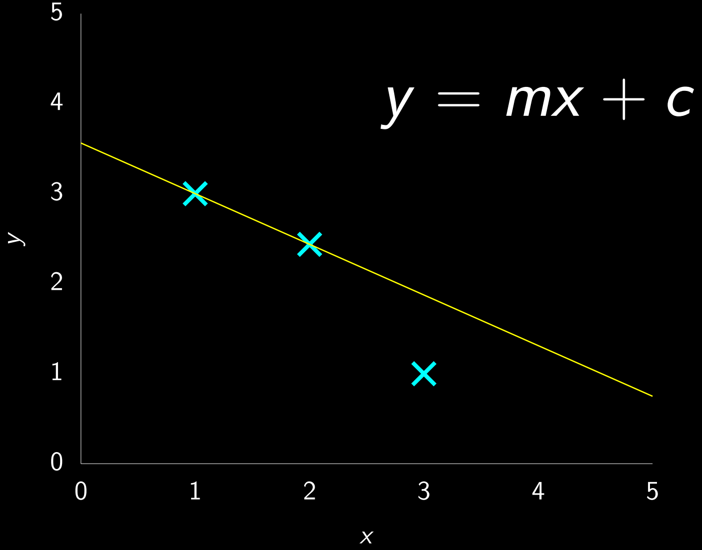
$$y = mx + c$$



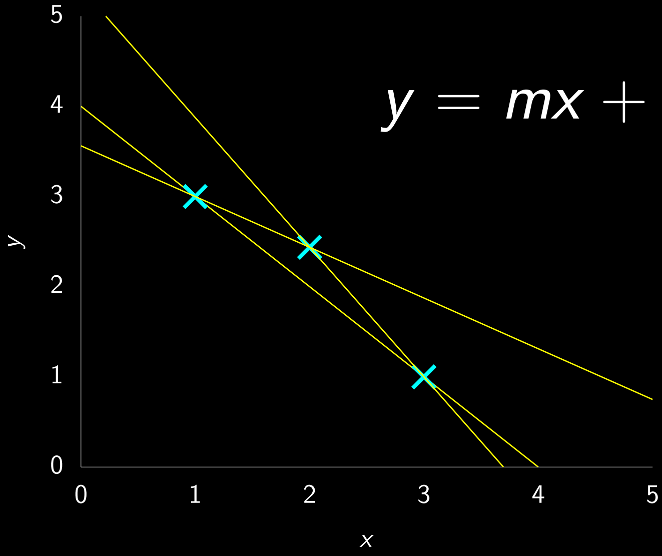

5
4
3
2
1
0
0
1
2
3
4
5
x
y
c
m
y = mx + c


$y = mx + c$
c
m
0
1
2
3
4
5
0
1
2
3
4
5
x
y


$y = mx + c$
5
4
3
2
1
0
0
1
2
3
4
5
y
x



$y = mx + c$
5
4
3
2
1
0
0
1
2
3
4
5
y
x



$$y = mx + c$$

$$y = mx + c$$

point 1: $x = 1$, $y = 3$

$$3 = m + c$$

point 2: $x = 3$, $y = 1$

$$1 = 3m + c$$

point 3: $x = 2$, $y = 2.5$

$$2.5 = 2m + c$$

$$y = mx + c + \epsilon$$

point 1: $x = 1$, $y = 3$

$$3 = m + c + \epsilon_1$$

point 2: $x = 3$, $y = 1$

$$1 = 3m + c + \epsilon_2$$

point 3: $x = 2$, $y = 2.5$

$$2.5 = 2m + c + \epsilon_3$$

What is Machine Learning?

Equipping Computers with Human Like Capabilities.

- Endow computers with the ability to "learn" from "data".
- Present data from sensors, the internet, experiments.
- Expect computer to make "sensible" decisions.
- Traditionally categorized as:

  Supervised learning: classification, regression.
  Unsupervised learning: dimensionality reduction, clustering.
  Reinforcement learning: learning from delayed feedback.
  Planning. Difficult stuff!

What is Machine Learning?

Equipping Computers with Human Like Capabilities.

- Endow computers with the ability to “learn” from “data”.
- Present data from sensors, the internet, experiments.
- Expect computer to make “sensible” decisions.
- Traditionally categorized as:

  Supervised learning: classification, regression.
  Unsupervised learning: dimensionality reduction, clustering.
  Reinforcement learning: learning from delayed feedback.
  Planning. Difficult stuff!

What is Machine Learning?

Equipping Computers with Human Like Capabilities.

- Endow computers with the ability to "learn" from "data".
- Present data from sensors, the internet, experiments.
- Expect computer to make "sensible" decisions.
- Traditionally categorized as:

  Supervised learning: classification, regression.
  Unsupervised learning: dimensionality reduction, clustering.
  Reinforcement learning: learning from delayed feedback.
  Planning. Difficult stuff!

What is Machine Learning?

Equipping Computers with Human Like Capabilities.

- Endow computers with the ability to “learn” from “data”.
- Present data from sensors, the internet, experiments.
- Expect computer to make “sensible” decisions.
- Traditionally categorized as:

  Supervised learning: classification, regression.

  Unsupervised learning: dimensionality reduction, clustering.

  Reinforcement learning: learning from delayed feedback.

  Planning. Difficult stuff!

What is Machine Learning?

Equipping Computers with Human Like Capabilities.

- Endow computers with the ability to "learn" from "data".
- Present data from sensors, the internet, experiments.
- Expect computer to make "sensible" decisions.
- Traditionally categorized as:

  Supervised learning: classification, regression.
  Unsupervised learning: dimensionality reduction, clustering.
  Reinforcement learning: learning from delayed feedback.
  Planning. Difficult stuff!

What is Machine Learning?

Equipping Computers with Human Like Capabilities.

- Endow computers with the ability to "learn" from "data".
- Present data from sensors, the internet, experiments.
- Expect computer to make "sensible" decisions.
- Traditionally categorized as:

  Supervised learning: classification, regression.
  Unsupervised learning: dimensionality reduction, clustering.
  Reinforcement learning: learning from delayed feedback.
  Planning. Difficult stuff!

Applications of Machine Learning

Handwriting Recognition : Recognising handwritten characters. For example LeNet `http://bit.ly/d26fwK`.

Ranking : Learning relative skills of on line game players, the TrueSkill system `http://research.microsoft.com/en-us/projects/trueskill/`.

Collaborative Filtering : Prediction of user preferences for items given purchase history. For example the Netflix Prize `http://www.netflixprize.com/`.

Internet Search : For example Ad Click Through rate prediction `http://bit.ly/a7XLH4`.

News Personalisation : For example Zite `http://www.zite.com/`.

Game Play Learning : For example, learning to play Go `http://bit.ly/cV77zM`.

History of Machine Learning (personal)

Rosenblatt to Vapnik

- Arises from the Connectionist movement in AI. `http://en.wikipedia.org/wiki/Connectionism`
- Early Connectionist research focused on models of the brain.

History of Machine Learning (personal)

Rosenblatt to Vapnik

- Arises from the Connectionist movement in AI.
  `http://en.wikipedia.org/wiki/Connectionism`
- Early Connectionist research focused on models of the brain.

Frank Rosenblatt's Perceptron

- Rosenblatt's perceptron (Rosenblatt, 1962) based on simple model of a neuron (McCulloch and Pitts, 1943) and a learning algorithm.

Figure: Frank Rosenblatt in 1950 (source: Cornell University Library)

Vladmir Vapnik's Statistical Learning Theory

- Later machine learning research focused on theoretical foundations of such models and their capacity to learn (Vapnik, 1998).

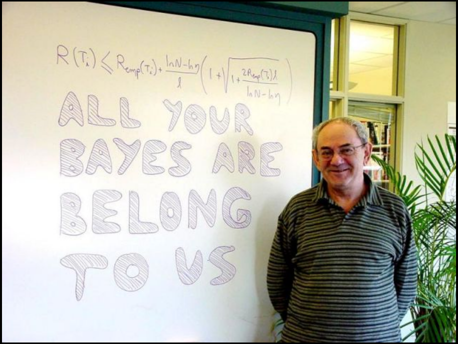


Figure: Vladimir Vapnik "All Your Bayes ..." (source `http://lecun.com/ex/fun/index.html`), see also `http://bit.ly/qfd2mU`.

Personal View

- Machine learning benefited greatly by incorporating ideas from psychology, but not being afraid to incorporate rigorous theory.

Machine Learning Today

An extension of statistics?

- Early machine learning viewed with scepticism by statisticians.
- Modern machine learning and statistics interact to both communities benefits.
- *Personal view*: statistics and machine learning are fundamentally different. Statistics aims to provide a human with the tools to analyze data. Machine learning wants to replace the human in the processing of data.

Machine Learning Today

An extension of statistics?

- Early machine learning viewed with scepticism by statisticians.
- Modern machine learning and statistics interact to both communities benefits.
- *Personal view*: statistics and machine learning are fundamentally different. Statistics aims to provide a human with the tools to analyze data. Machine learning wants to replace the human in the processing of data.

Machine Learning Today

An extension of statistics?

- Early machine learning viewed with scepticism by statisticians.
- Modern machine learning and statistics interact to both communities benefits.
- *Personal view*: statistics and machine learning are fundamentally different. Statistics aims to provide a human with the tools to analyze data. Machine learning wants to replace the human in the processing of data.

Machine Learning Today

Mathematics and Bumblebees

- For the moment the two overlap strongly. But they are not the same field!
- Machine learning also has overlap with Cognitive Science.
- Mathematical formalisms of a problem are helpful, but they can hide facts: i.e. the fallacy that "aerodynamically a bumble bee can't fly". Clearly a limitation of the model rather than fact.
- Mathematical foundations are still very important though: they help us understand the capabilities of our algorithms.
- But we mustn't restrict our ambitions to the limitations of current mathematical formalisms. That is where humans give inspiration.

Machine Learning Today

Mathematics and Bumblebees

- For the moment the two overlap strongly. But they are not the same field!
- Machine learning also has overlap with Cognitive Science.
- Mathematical formalisms of a problem are helpful, but they can hide facts: i.e. the fallacy that "aerodynamically a bumble bee can't fly". Clearly a limitation of the model rather than fact.
- Mathematical foundations are still very important though: they help us understand the capabilities of our algorithms.
- But we mustn't restrict our ambitions to the limitations of current mathematical formalisms. That is where humans give inspiration.

Machine Learning Today

Mathematics and Bumblebees

- For the moment the two overlap strongly. But they are not the same field!
- Machine learning also has overlap with Cognitive Science.
- Mathematical formalisms of a problem are helpful, but they can hide facts: i.e. the fallacy that "aerodynamically a bumble bee can't fly". Clearly a limitation of the model rather than fact.
- Mathematical foundations are still very important though: they help us understand the capabilities of our algorithms.
- But we mustn't restrict our ambitions to the limitations of current mathematical formalisms. That is where humans give inspiration.

Machine Learning Today

Mathematics and Bumblebees

- For the moment the two overlap strongly. But they are not the same field!
- Machine learning also has overlap with Cognitive Science.
- Mathematical formalisms of a problem are helpful, but they can hide facts: i.e. the fallacy that "aerodynamically a bumble bee can't fly". Clearly a limitation of the model rather than fact.
- Mathematical foundations are still very important though: they help us understand the capabilities of our algorithms.
- But we mustn't restrict our ambitions to the limitations of current mathematical formalisms. That is where humans give inspiration.

Machine Learning Today

Mathematics and Bumblebees

- For the moment the two overlap strongly. But they are not the same field!
- Machine learning also has overlap with Cognitive Science.
- Mathematical formalisms of a problem are helpful, but they can hide facts: i.e. the fallacy that "aerodynamically a bumble bee can't fly". Clearly a limitation of the model rather than fact.
- Mathematical foundations are still very important though: they help us understand the capabilities of our algorithms.
- But we mustn't restrict our ambitions to the limitations of current mathematical formalisms. That is where humans give inspiration.

Statistics

What's in a Name?

- Early statistics had great success with the idea of statistical proof.

  Question: I computed the mean of these two tables of numbers (a statistic). They are different. Does this "prove" anything?

  Answer: it depends on how the numbers are generated, how many there are and how big the difference. Randomization is important.

- Hypothesis testing: questions you can ask about your data are quite limiting.
- This can have the affect of limiting science too.
- Many successes: crop fertilization, clinical trials, brewing, polling.
- Many open questions: e.g. causality.

Statistics

What's in a Name?

- Early statistics had great success with the idea of statistical proof.

  Question: I computed the mean of these two tables of numbers (a statistic). They are different. Does this "prove" anything?

  Answer: it depends on how the numbers are generated, how many there are and how big the difference. Randomization is important.

- Hypothesis testing: questions you can ask about your data are quite limiting.
- This can have the affect of limiting science too.
- Many successes: crop fertilization, clinical trials, brewing, polling.
- Many open questions: e.g. causality.

Statistics

What's in a Name?

- Early statistics had great success with the idea of statistical proof.

  Question: I computed the mean of these two tables of numbers (a statistic). They are different. Does this "prove" anything?

  Answer: it depends on how the numbers are generated, how many there are and how big the difference. Randomization is important.

- Hypothesis testing: questions you can ask about your data are quite limiting.
- This can have the affect of limiting science too.
- Many successes: crop fertilization, clinical trials, brewing, polling.
- Many open questions: e.g. causality.

Statistics

What's in a Name?

- Early statistics had great success with the idea of statistical proof.

  Question: I computed the mean of these two tables of numbers (a statistic). They are different. Does this "prove" anything?

  Answer: it depends on how the numbers are generated, how many there are and how big the difference. Randomization is important.

- Hypothesis testing: questions you can ask about your data are quite limiting.
- This can have the affect of limiting science too.
- Many successes: crop fertilization, clinical trials, brewing, polling.
- Many open questions: e.g. causality.

Statistics

What's in a Name?

- Early statistics had great success with the idea of statistical proof.

  Question: I computed the mean of these two tables of numbers (a statistic). They are different. Does this "prove" anything?

  Answer: it depends on how the numbers are generated, how many there are and how big the difference. Randomization is important.

- Hypothesis testing: questions you can ask about your data are quite limiting.
- This can have the affect of limiting science too.
- Many successes: crop fertilization, clinical trials, brewing, polling.
- Many open questions: e.g. causality.

Statistics

What's in a Name?

- Early statistics had great success with the idea of statistical proof.

  Question: I computed the mean of these two tables of numbers (a statistic). They are different. Does this "prove" anything?

  Answer: it depends on how the numbers are generated, how many there are and how big the difference. Randomization is important.

- Hypothesis testing: questions you can ask about your data are quite limiting.
- This can have the affect of limiting science too.
- Many successes: crop fertilization, clinical trials, brewing, polling.
- Many open questions: e.g. causality.

Early 20th Century Statistics

- Many statisticians were Edwardian English gentleman.

Figure: William Sealy Gosset in 1908

Statistics and Machine Learning

Statisticians want to turn humans into computers.
Machine learners want to turn computers into humans.
We meet somewhere in the middle.

NDL 2012/06/16

Statistics

- Cricket and Baseball are two games with a lot of “statistics”.
- The study of the meaning behind these numbers is “mathematical statistics” often abbreviated to “statistics”.

Statistics

- Cricket and Baseball are two games with a lot of “statistics”.
- The study of the meaning behind these numbers is “mathematical statistics” often abbreviated to “statistics”.

Machine Learning and Probability

- The world is an *uncertain* place.

  Epistemic uncertainty: uncertainty arising through lack of knowledge. (What colour socks is that person wearing?)

  Aleatoric uncertainty: uncertainty arising through an underlying stochastic system. (Where will a sheet of paper fall if I drop it?)

Machine Learning and Probability

- The world is an *uncertain* place.

  Epistemic uncertainty: uncertainty arising through lack of knowledge. (What colour socks is that person wearing?)

  Aleatoric uncertainty: uncertainty arising through an underlying stochastic system. (Where will a sheet of paper fall if I drop it?)

Machine Learning and Probability

- The world is an *uncertain* place.

  Epistemic uncertainty: uncertainty arising through lack of knowledge. (What colour socks is that person wearing?)

  Aleatoric uncertainty: uncertainty arising through an underlying stochastic system. (Where will a sheet of paper fall if I drop it?)

Probability: A Framework to Characterise Uncertainty

- We need a framework to characterise the uncertainty.
- In this course we make use of probability theory to characterise uncertainty.

Probability: A Framework to Characterise Uncertainty

- We need a framework to characterise the uncertainty.
- In this course we make use of probability theory to characterise uncertainty.

Richard Price

- Welsh philosopher and essay writer.
- Edited **Thomas Bayes**'s essay which contained foundations of Bayesian philosophy.

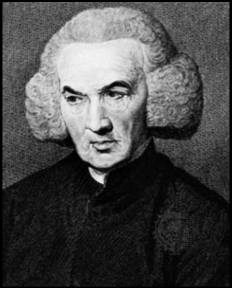


Figure: Richard Price, 1723–1791. (source Wikipedia)

Laplace

- French Mathematician and Astronomer.

Figure: Pierre-Simon Laplace, 1749–1827. (source Wikipedia)

Styles of Machine Learning

Background: interpolation is easy, extrapolation is hard

- Urs Hölzle keynote talk at NIPS 2005.
    - Emphasis on massive data sets.
    - Let the data do the work—more data, less extrapolation.
- Alternative paradigm:
    - Very scarce data: computational biology, human motion.
    - How to generalize from scarce data?
    - Need to include more assumptions about the data (e.g. invariances).

General Approach

Broadly Speaking: Two approaches to modeling

data modeling	*mechanistic modeling*

General Approach

Broadly Speaking: Two approaches to modeling

data modeling	*mechanistic modeling*
let the data "speak"	

General Approach

Broadly Speaking: Two approaches to modeling

data modeling	*mechanistic modeling*
let the data “speak”	impose physical laws

General Approach

Broadly Speaking: Two approaches to modeling

data modeling	*mechanistic modeling*
let the data "speak"	impose physical laws
data driven	

General Approach

Broadly Speaking: Two approaches to modeling

data modeling	*mechanistic modeling*
let the data "speak"	impose physical laws
data driven	knowledge driven

General Approach

Broadly Speaking: Two approaches to modeling

data modeling	*mechanistic modeling*
let the data "speak"	impose physical laws
data driven	knowledge driven
adaptive models	

General Approach

Broadly Speaking: Two approaches to modeling

data modeling	*mechanistic modeling*
let the data "speak"	impose physical laws
data driven	knowledge driven
adaptive models	differential equations

General Approach

Broadly Speaking: Two approaches to modeling

data modeling	*mechanistic modeling*
let the data "speak"	impose physical laws
data driven	knowledge driven
adaptive models	differential equations
digit recognition	

General Approach

Broadly Speaking: Two approaches to modeling

data modeling	*mechanistic modeling*
let the data "speak"	impose physical laws
data driven	knowledge driven
adaptive models	differential equations
digit recognition	climate, weather models

General Approach

Broadly Speaking: Two approaches to modeling

data modeling	*mechanistic modeling*
let the data "speak"	impose physical laws
data driven	knowledge driven
adaptive models	differential equations
digit recognition	climate, weather models

Weakly Mechanistic

General Approach

Broadly Speaking: Two approaches to modeling

data modeling	*mechanistic modeling*
let the data "speak"	impose physical laws
data driven	knowledge driven
adaptive models	differential equations
digit recognition	climate, weather models

Weakly Mechanistic

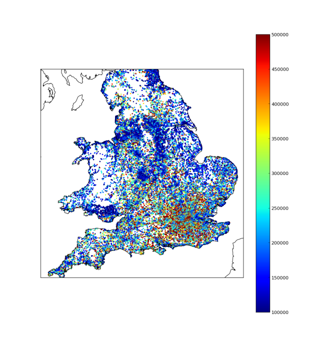
Strongly Mechanistic

Weakly Mechanistic vs Strongly Mechanistic

- Underlying data modeling techniques there are *weakly mechanistic* principles (e.g. smoothness).
- In physics the models are typically *strongly mechanistic*.
- In principle we expect a range of models which vary in the strength of their mechanistic assumptions.
- This work is one part of that spectrum: add further mechanistic ideas to weakly mechanistic models.

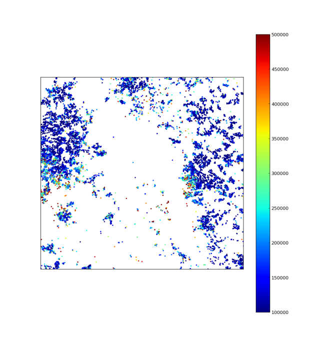
What's Changed (Changing) for Medicine?

- Modern data availability.



What's Changed (Changing) for Medicine?

- Modern data availability.



What's Changed (Changing) for Medicine?

- Google: patient data ...

What's Changed (Changing) for Medicine?

- The Red Flag Analogy.
- ... but why I work in Medical data ...

What's Changed (Changing) for Medicine?

- Genotyping.
- Epigenotyping.
- Transcriptome: detailed characterization of phenotype.
  - Self-organizing-stratifications of data.
- Automatic data curation: from curated data to curation of publicly available data.
- Patient Access: `http://www.patient.co.uk/patient-access.asp`
- Open Data: `http://www.openstreetmap.org/?lat=53.38086&lon=-1.48545&zoom=17&layers=M`.
- Tescos and Facebook.

References I

Monatliche Correspondenz zur beförderung der Erd- und Himmels-kunde. Number v. 4. Beckerische Buchhandlung., 1801. [URL].

C. F. Gauss. Astronomische untersuchungen und rechnungen vornehmlich ÃŒber die ceres ferdinandea, 1802. Nachlass Gauss, Handbuch 4, Bl. 1.

W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943. Reprinted in **?**.

F. Rosenblatt. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan, 1962.

V. N. Vapnik. *Statistical Learning Theory*. John Wiley and Sons, New York, 1998.