Machine Learning Motivation

MLAI Lecture 1

Neil D. Lawrence

Department of Computer Science
Sheffield University

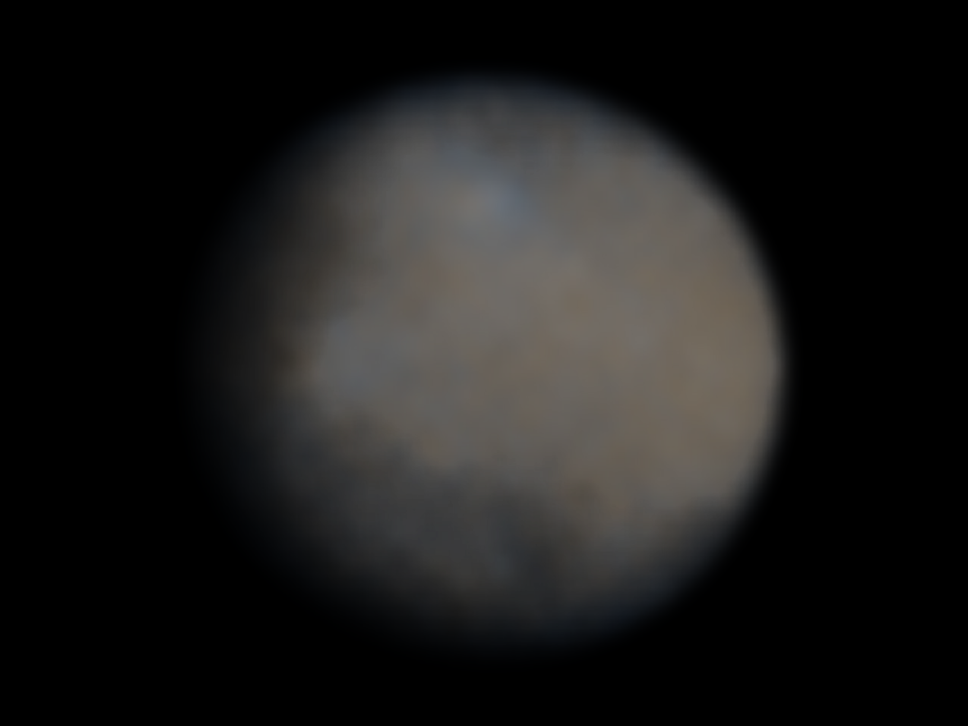
24th September 2012

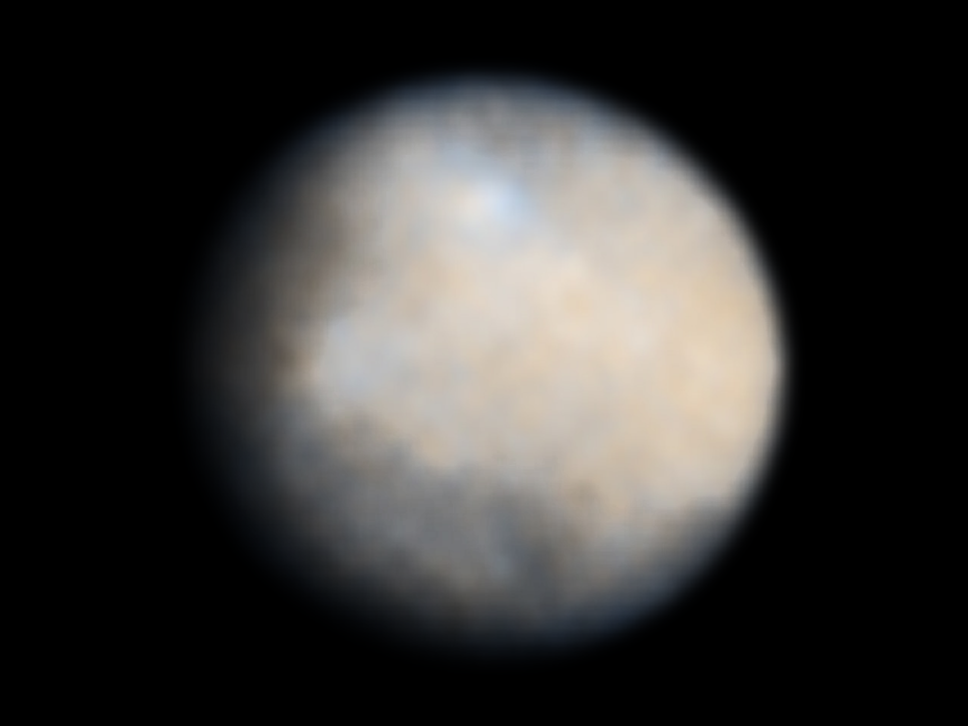
Outline

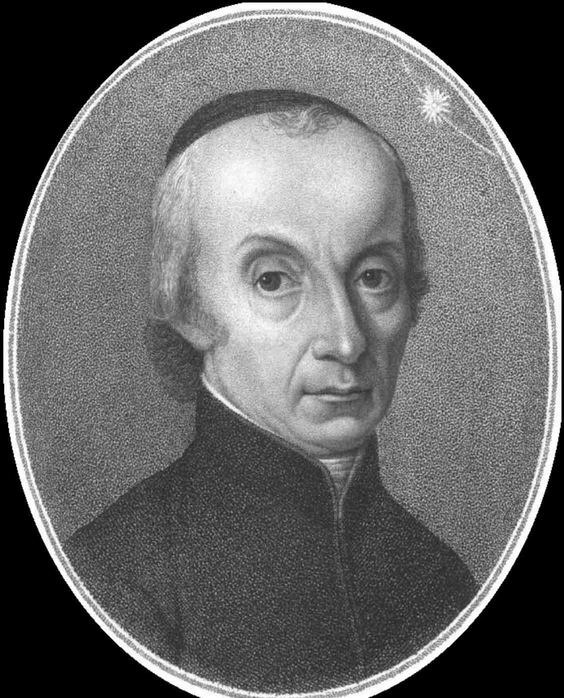
Motivation

Course Text

Basic Probability







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 27,8	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[illegible]	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[illegible]	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,[illegible]	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[illegible]	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:

Sonnenferne . .	326° 27′ 38″		Hieraus;	
☊	81 0 44		größte Mittelp. Gleichung	9° 27′ 41″
Neigung . .	10 36 57		tägl. mittlere helioc. tropiſche Beweg.	770,″914
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*

Excentricität . . 0,0825017	tropische Beweg. 770, 914
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

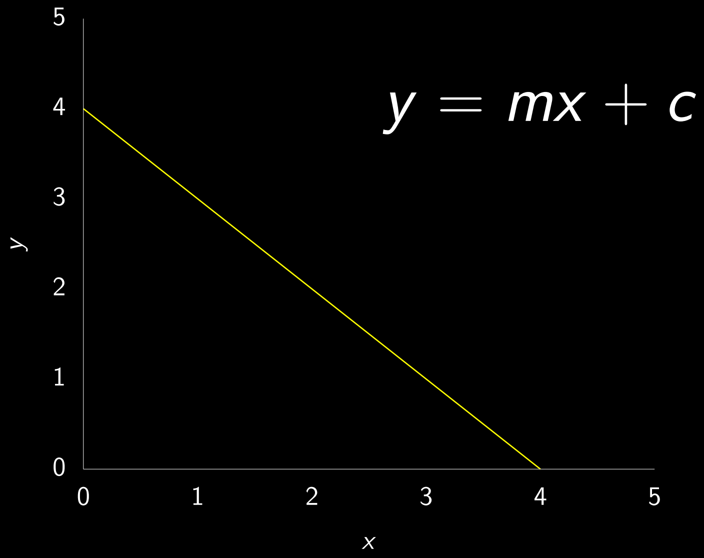
CERERE FERDINANDEA

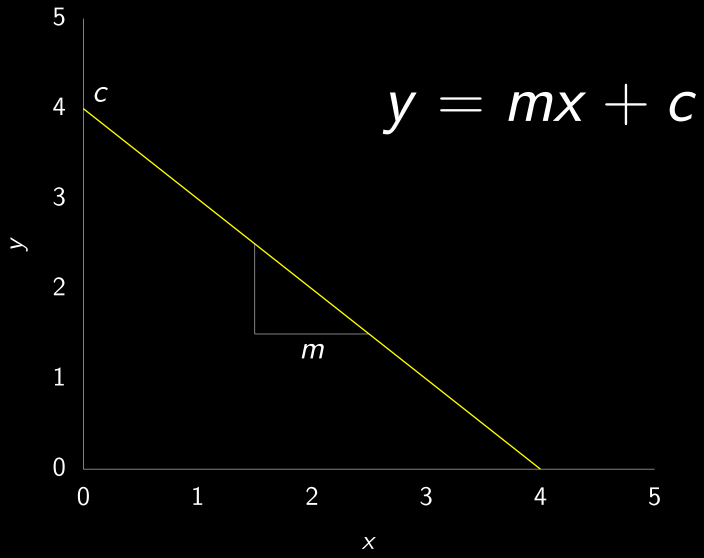
OTTAVO TRA I PRIMARJ DEL NOSTRO SISTEMA
SOLARE.

PALERMO
1802

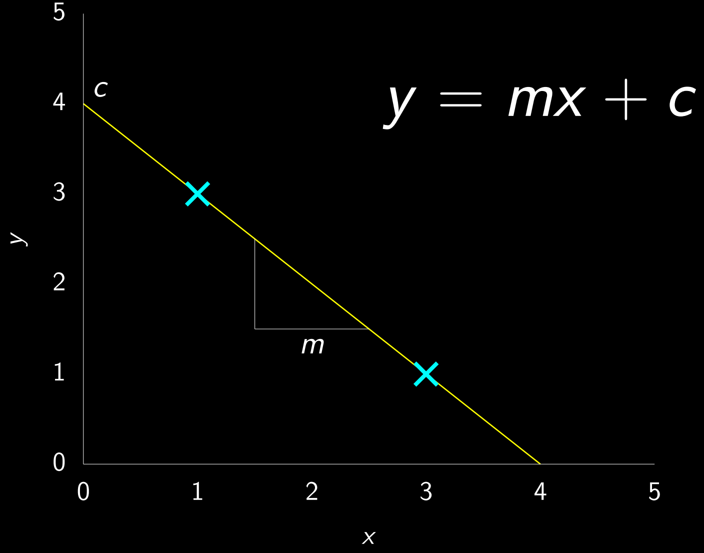
NELLA STAMPERIA REALE.

$$y = mx + c$$

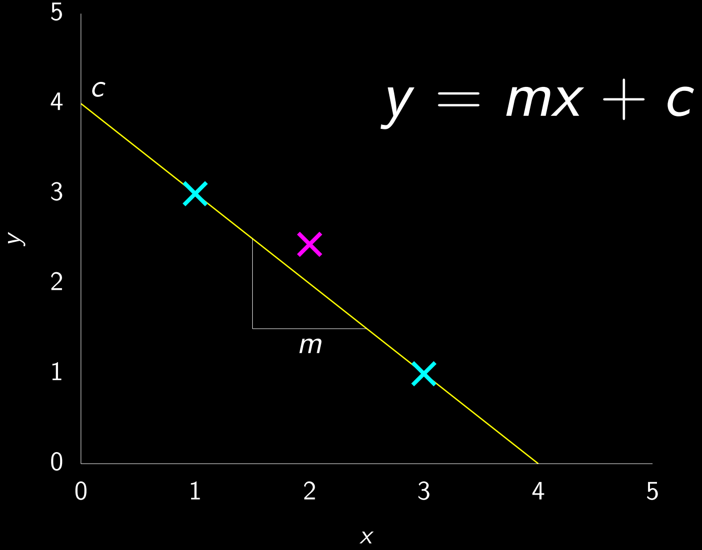

$y = mx + c$
5
4
3
2
1
0
y
0
1
2
3
4
5
x



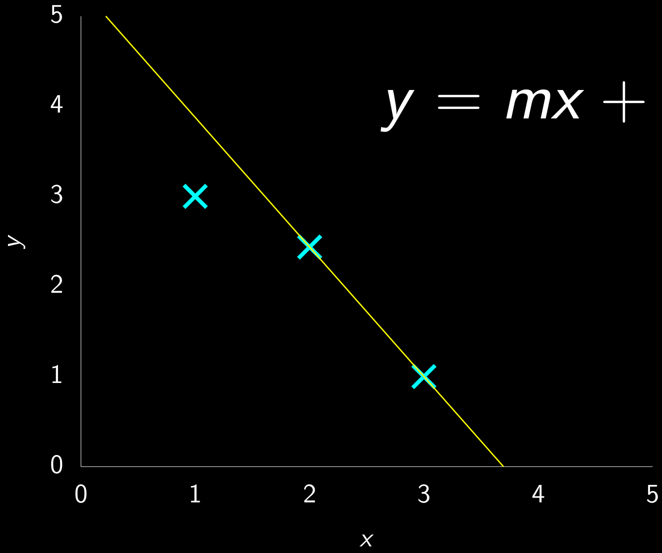
$y = mx + c$
c
m
y
x
0
1
2
3
4
5



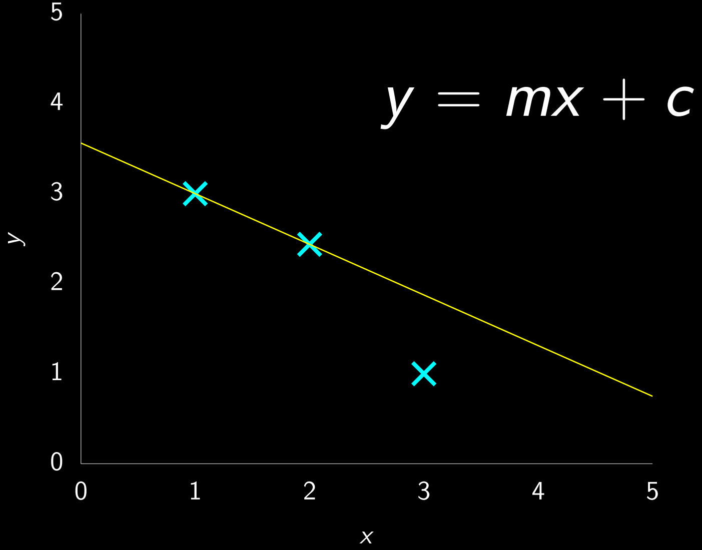
$y = mx + c$
c
m
0
1
2
3
4
5
x
y

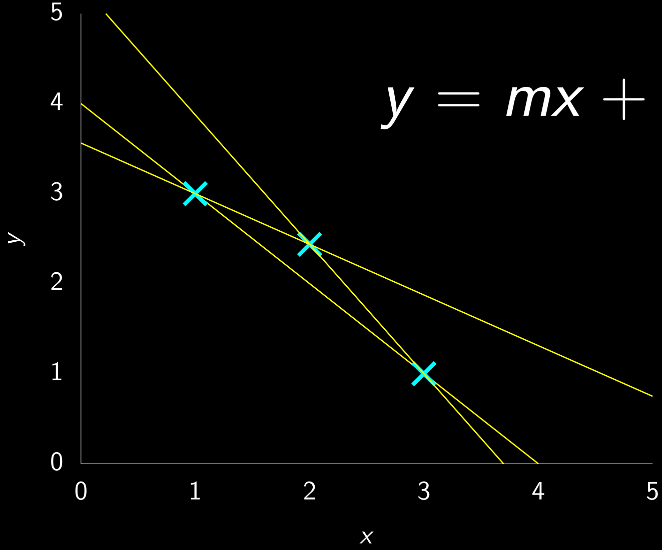


$y = mx + c$
c
m
y
x
0
1
2
3
4
5



$$y = mx + c$$


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



$$y = mx + c$$

$$y = mx + c$$

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

$$3 = m + c$$

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

$$1 = 3m + c$$

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

$$2.5 = 2m + c$$

riens. L'opinion contraire est une illusion de l'esprit qui, perdant de vue les raisons fugitives du choix de la volonté dans les choses indifférentes, se persuade qu'elle s'est déterminée d'elle-même et sans motifs.

Nous devons donc envisager l'état présent de l'univers, comme l'effet de son état antérieur, et comme la cause de celui qui va suivre. Une intelligence qui, pour un instant donné, connaîtrait toutes les forces dont la nature est animée, et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'analyse, embrasserait dans la même formule les mouvemens des plus grands corps de l'univers et ceux du plus léger atome : rien ne serait incertain pour elle, et l'avenir comme le passé, serait présent à ses yeux. L'esprit humain offre, dans la perfection qu'il a su donner à l'Astronomie, une faible esquisse de cette intelligence. Ses découvertes en Mécanique et en Géométrie, jointes à celle de la pesanteur universelle, l'ont mis à portée de comprendre dans les mêmes expressions analytiques, les états passés et futurs du système du monde. En appliquant la même méthode à quelques autres objets de ses connaissances, il est parvenu à ramener à des lois générales, les phénomènes observés, et à prévoir ceux que des circonstances données doivent faire éclore. Tous ces efforts dans la recherche de la vérité, tendent à le rapprocher sans cesse de l'intelligence que nous venons de concevoir, mais dont il restera toujours infiniment éloigné. Cette tendance propre à l'espèce humaine, est ce qui la rend supérieure aux animaux; et ses progrès en ce genre, distinguent les nations et les siècles, et font leur véritable gloire.

Rappelons-nous qu'autrefois, et à une époque qui height: "The day will come when, by study pursued through several ages, the things now concealed will appear with evidence; and posterity will be astonished that truths so clear had escaped us." Clairaut then undertook to submit to analysis the perturbations which the comet had experienced by the action of the two great planets, Jupiter and Saturn; after immense calculations he fixed its next passage at the perihelion toward the beginning of April, 1759, which was actually verified by observation. The regularity which astronomy shows us in the movements of the comets doubtless exists also in all phenomena.

The curve described by a simple molecule of air or vapor is regulated in a manner just as certain as the planetary orbits; the only difference between them is that which comes from our ignorance.

Probability is relative, in part to this ignorance, in part to our knowledge. We know that of three or a greater number of events a single one ought to occur; but nothing induces us to believe that one of them will occur rather than the others. In this state of indecision it is impossible for us to announce their occurrence with certainty. It is, however, probable that one of these events, chosen at will, will not occur because we see several cases equally possible which exclude its occurrence, while only a single one favors it.

The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of shows us in the movements of the comets doubtless exists also in all phenomena.

The curve described by a simple molecule of air or vapor is regulated in a manner just as certain as the planetary orbits; the only difference between them is that which comes from our ignorance.

Probability is relative, in part to this ignorance, in part to our knowledge. We know that of three or a other, we say that its choice is an effect without a cause. It is then, says Leibnitz, the blind chance of the Epicureans. The contrary opinion is an illusion of the mind, which, losing sight of the evasive reasons of the choice of the will in indifferent things, believes that choice is determined of itself and without motives.

We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis—it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes. The human mind offers, in the perfection which it has been able to give to astronomy, a feeble idea of this intelligence. Its discoveries in mechanics and geometry, added to that of universal gravity, have enabled it to comprehend in the same analytical expressions the past and future states of the system of the world. Applying the same method to some other objects of its knowledge, it has succeeded in referring to general laws observed phenomena and in foreseeing those which given circumstances ought to produce. All these efforts in the search for truth tend to lead it back continually to the vast intelligence which we have just mentioned, but from which it will always remain infinitely removed. This tendency, peculiar to the human race, is that which renders it superior to animals; and their progress

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

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).

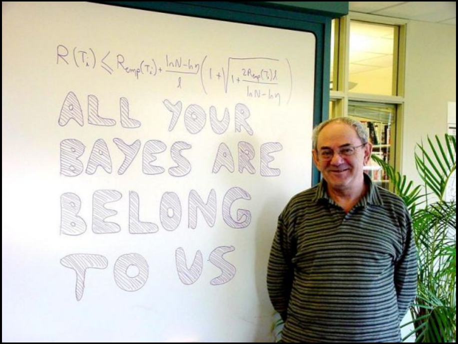


Figure: Vladimir Vapnik "All Your Bayes ..." (source `http://lecun.com/ex/fun/index.html`), see also `http://bit.ly/qfd2mU`.

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.

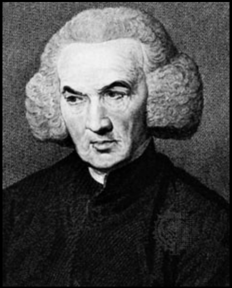


Figure: Richard Price, 1723–1791. (source Wikipedia)

Laplace

- French Mathematician and Astronomer.

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

Outline

Motivation

Course Text

Basic Probability


PATTERN RECOGNITION
AND MACHINE LEARNING
CHRISTOPHER M. BISHOP

Outline

Motivation

Course Text

Basic Probability

Probability Review I

- We are interested in trials which result in two random variables, X and Y, each of which has an 'outcome' denoted by x or y.
- We summarise the notation and terminology for these distributions in the following table.

Terminology	Notation	Description
Joint Probability	$P(X = x, Y = y)$	'The probability that $X = x$ *and* $Y = y$'
Marginal Probability	$P(X = x)$	'The probability that $X = x$ *regardless of* Y'
Conditional Probability	$P(X = x \mid Y = y)$	'The probability that $X = x$ *given that* $Y = y$'

Table: The different basic probability distributions.

A Pictorial Definition of Probability

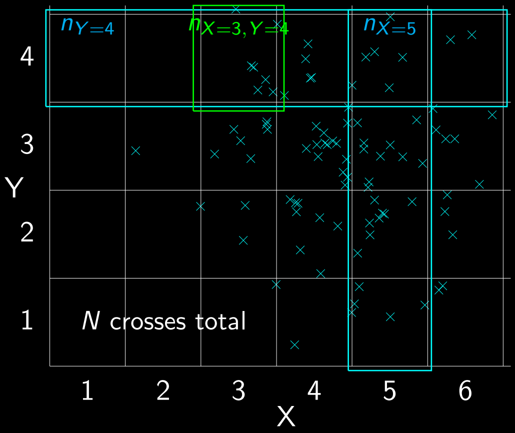


Figure: Representation of joint and conditional probabilities.

Different Distributions

Terminology	Definition	Notation
Joint Probability	$\lim_{N\to\infty} \frac{n_{X=3,Y=4}}{N}$	$P(X=3, Y=4)$
Marginal Probability	$\lim_{N\to\infty} \frac{n_{X=5}}{N}$	$P(X=5)$
Conditional Probability	$\lim_{N\to\infty} \frac{n_{X=3,Y=4}}{n_{Y=4}}$	$P(X=3 \mid Y=4)$

Table: Definition of probability distributions.

Notational Details

- Typically we should write out $P(X = x, Y = y)$.
- In practice, we often use $P(x, y)$.
- This looks very much like we might write a multivariate function, *e.g.* $f(x, y) = \frac{x}{y}$.
  - For a multivariate function though, $f(x, y) \neq f(y, x)$.
  - However $P(x, y) = P(y, x)$ because $P(X = x, Y = y) = P(Y = y, X = x)$.
- We now quickly review the 'rules of probability'.

Normalization

All distributions are normalized. This is clear from the fact that $\sum_x n_x = N$, which gives

$$\sum_x P(x) = \frac{\sum_x n_x}{N} = \frac{N}{N} = 1.$$

A similar result can be derived for the marginal and conditional distributions.

The Sum Rule

Ignoring the limit in our definitions:

- The marginal probability $P(y)$ is $\frac{n_y}{N}$ (ignoring the limit).
- The joint distribution $P(x, y)$ is $\frac{n_{x,y}}{N}$.
- $n_y = \sum_x n_{x,y}$ so
$$\frac{n_y}{N} = \sum_x \frac{n_{x,y}}{N},$$
in other words
$$P(y) = \sum_x P(x, y).$$
This is known as the sum rule of probability.

The Product Rule

- $P(x|y)$ is
$$\frac{n_{x,y}}{n_y}.$$
- $P(x,y)$ is
$$\frac{n_{x,y}}{N} = \frac{n_{x,y}}{n_y}\frac{n_y}{N}$$
or in other words
$$P(x,y) = P(x|y)\,P(y)\,.$$
This is known as the product rule of probability.

Bayes' Rule

- From the product rule,

$$P(y, x) = P(x, y) = P(x|y)\,P(y)\,,$$

so

$$P(y|x)\,P(x) = P(x|y)\,P(y)$$

which leads to Bayes' rule,

$$P(y|x) = \frac{P(x|y)\,P(y)}{P(x)}.$$

Bayes' Theorem Example

- There are two barrels in front of you. Barrel One contains 20 apples and 4 oranges. Barrel Two other contains 4 apples and 8 oranges. You choose a barrel randomly and select a fruit. It is an apple. What is the probability that the barrel was Barrel One?

Bayes' Theorem Example: Answer I

- We are given that:

$$
\begin{aligned}
P(\mathsf{F} = \mathsf{A} | \mathsf{B} = 1) =& 20/24 \\
P(\mathsf{F} = \mathsf{A} | \mathsf{B} = 2) =& 5/12 \\
P(\mathsf{B} = 1) =& 0.5 \\
P(\mathsf{B} = 2) =& 0.5
\end{aligned}
$$

Bayes' Theorem Example: Answer II

- We use the sum rule to compute:

$$\begin{aligned} P(\mathsf{F}=\mathsf{A}) =& P(\mathsf{F}=\mathsf{A}|\mathsf{B}=1)P(\mathsf{B}=1) \\ &+ P(\mathsf{F}=\mathsf{A}|\mathsf{B}=2)P(\mathsf{B}=2) \\ =& 20/24 \times 0.5 + 4/12 \times 0.5 = 7/12 \end{aligned}$$

- And Bayes' theorem tells us that:

$$\begin{aligned} P(\mathsf{B}=1|\mathsf{F}=\mathsf{A}) =& \frac{P(\mathsf{F}=\mathsf{A}|\mathsf{B}=1)P(\mathsf{B}=1)}{P(\mathsf{F}=\mathsf{A})} \\ =& \frac{20/24 \times 0.5}{7/12} = 5/7 \end{aligned}$$

Bayes' Theorem Example: Answer II

- We use the sum rule to compute:

$$\begin{aligned} P(\mathsf{F}=\mathsf{A}) =& P(\mathsf{F}=\mathsf{A}|\mathsf{B}=1)P(\mathsf{B}=1) \\ &+ P(\mathsf{F}=\mathsf{A}|\mathsf{B}=2)P(\mathsf{B}=2) \\ =& 20/24 \times 0.5 + 4/12 \times 0.5 = 7/12 \end{aligned}$$

- And Bayes' theorem tells us that:

$$\begin{aligned} P(\mathsf{B}=1|\mathsf{F}=\mathsf{A}) =& \frac{P(\mathsf{F}=\mathsf{A}|\mathsf{B}=1)P(\mathsf{B}=1)}{P(\mathsf{F}=\mathsf{A})} \\ =& \frac{20/24 \times 0.5}{7/12} = 5/7 \end{aligned}$$

Reading & Exercises

Before Friday's lecture:

- Read and *understand* Bishop on probability distributions: page 12–17 (Section 1.2).
- Complete Exercise 1.3 in Bishop.

References I

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

J. A. Anderson and E. Rosenfeld, editors. *Neurocomputing: Foundations of Research*, Cambridge, MA, 1988. MIT Press.

C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006. [Google Books] .

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 Anderson and Rosenfeld (1988).

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.