UNIVERSITY OF CAMBRIDGE

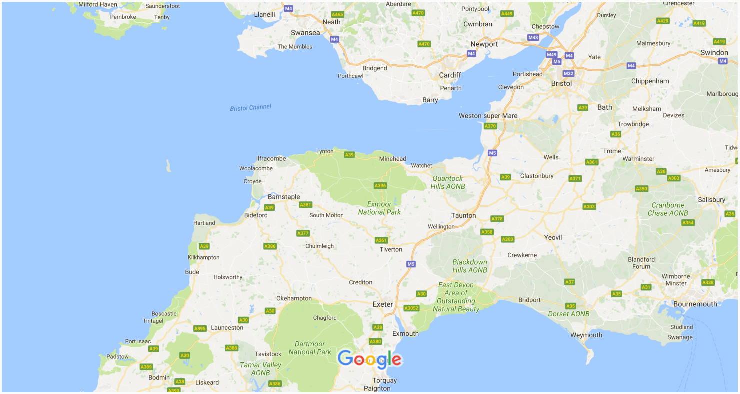
Machine Learning and the Scientific Principle

Carl Henrik Ek - che29@cam.ac.uk

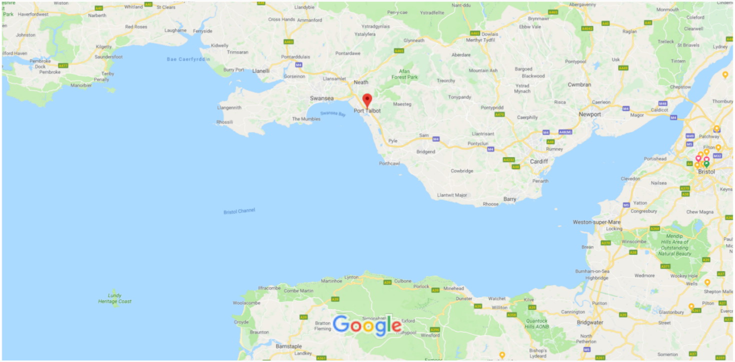
2nd of February, 2021

http://carlhenrik.com

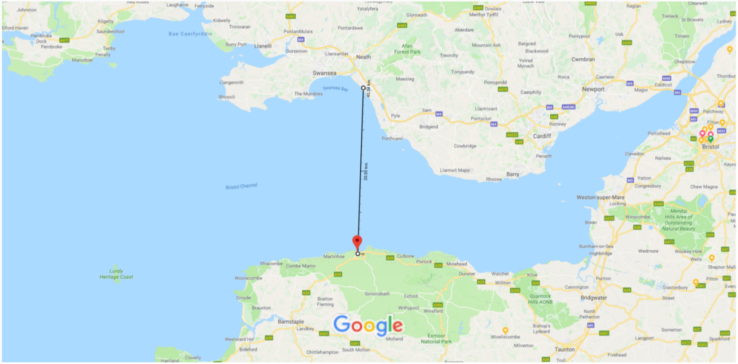
Cyder Country

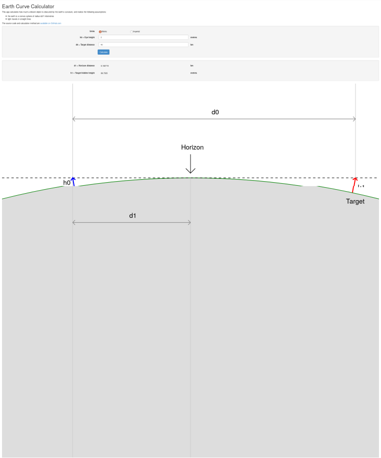


My Dream!


Swansea
Port Talbot
Neath
Llanelli
Cardiff
Newport
Cwmbran
Barry
Bristol
Weston-super-Mare
Bridgwater
Barnstaple
Bristol Channel
Bae Caerfyrddin
Swansea Bay
The Mumbles
Porthcawl
Bridgend
Pyle
Maesteg
Caerphilly
Pontypridd
Penarth
Portishead
Clevedon
Nailsea
Minehead
Lynton
Ilfracombe
Lundy Heritage Coast
Mendip Hills Area of Outstanding Natural Beauty
Quantock Hills AONB
Burnham-on-Sea
Pembroke
Tenby
Chepstow
Google

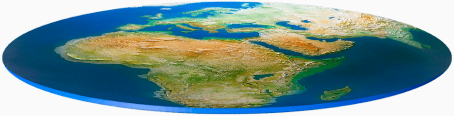

Swansea
Neath
Port Talbot
Llanelli
The Mumbles
Swansea Bay
Bae Caerfyrddin
Cardiff
Newport
Cwmbran
Bridgend
Barry
Penarth
Bristol
Portishead
Clevedon
Nailsea
Weston-super-Mare
Bristol Channel
Lundy
Heritage Coast
Lynton
Martinhoe
Ilfracombe
Combe Martin
Woolacombe
Porlock
Minehead
Dunster
Watchet
Barnstaple
Bideford
Exmoor
National Park
Bridgwater
Taunton
Quantock
Hills AONB
Mendip
Hills Area of
Outstanding
Natural Beauty
Tenby
Pembroke
Google


Swansea
Neath
Llanelli
Swansea Bay
The Mumbles
Bridgend
Cardiff
Newport
Cwmbran
Barry
Penarth
Cowbridge
Porthcawl
Bristol Channel
20.00 km
Lundy
Heritage Coast
Bristol
Portishead
Clevedon
Weston-super-Mare
Burnham-on-Sea
Bridgwater
Taunton
Minehead
Porlock
Ilfracombe
Combe Martin
Woolacombe
Braunton
Barnstaple
Bideford
Exmoor
National Park
Glastonbury
Wells
Google



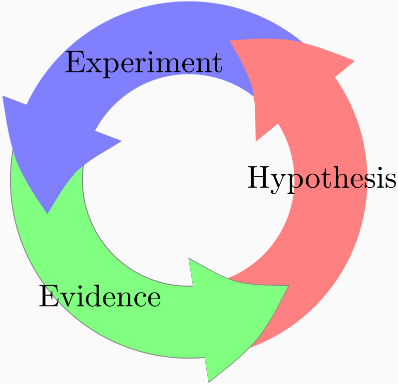
Distance to horizon $6.2km$

Hidden height $125.6m$

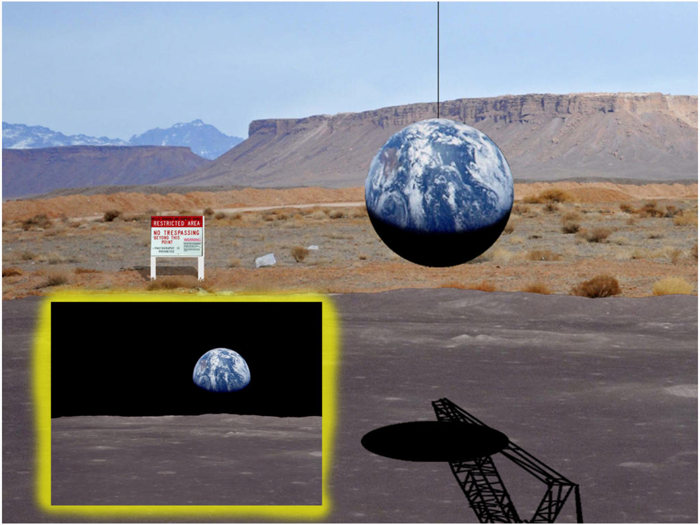


Well it kinda looks like a ball from the moon

Why does this not make sense?



Or not


RESTRICTED AREA
NO TRESPASSING
BEYOND THIS POINT
WARNING
PHOTOGRAPHY IS PROHIBITED

Laplace Demon

All these efforts in the search for truth tend to lead the human mind back continually to the vast intelligence which we have just mentioned, but from which it will always remain infinitely removed.
– Laplace Laplace, 1814a


OCCAM'S
RAZOR
... the simplest explanation is usually
the correct one.

Today

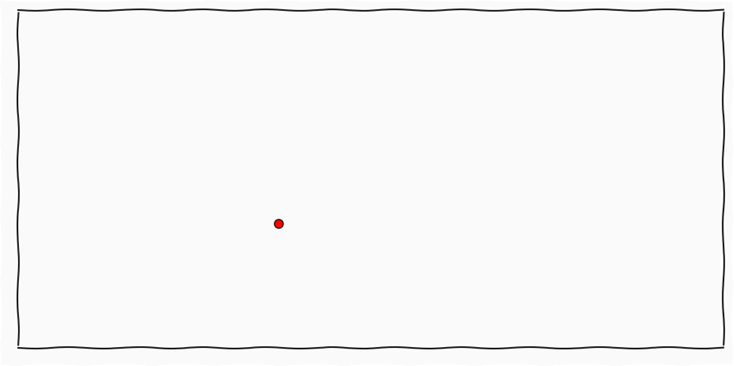
Where does machine learning fit into the scientific workflow?

1. How do we implement the scientific principle?
2. How do we implement Occam's razor?

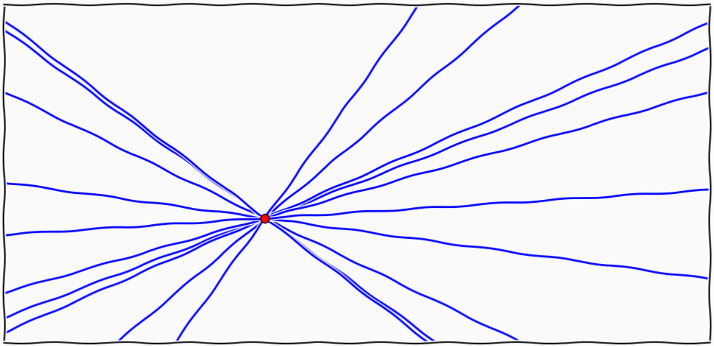
Machine Learning

Linear Regression

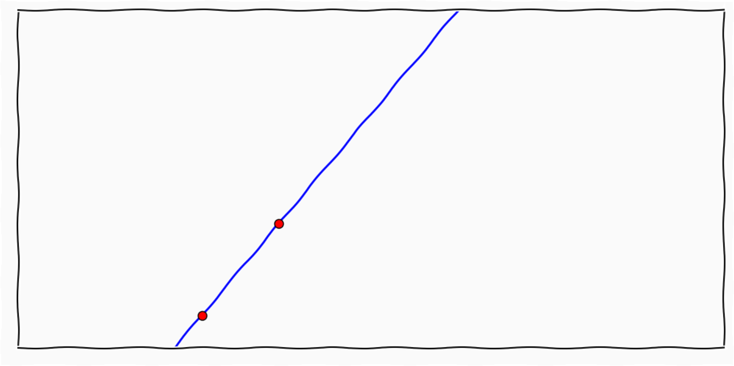
Linear Regression



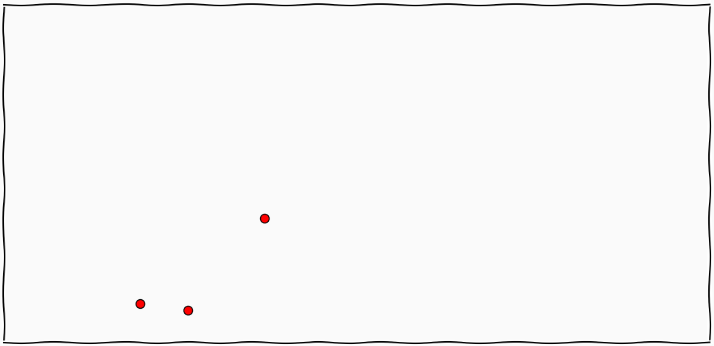
Linear Regression



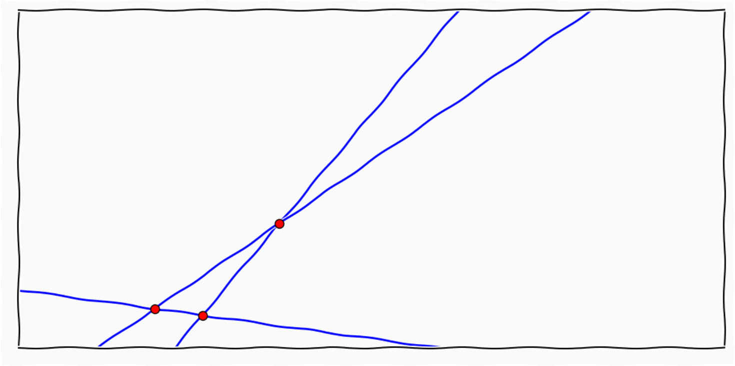
Linear Regression



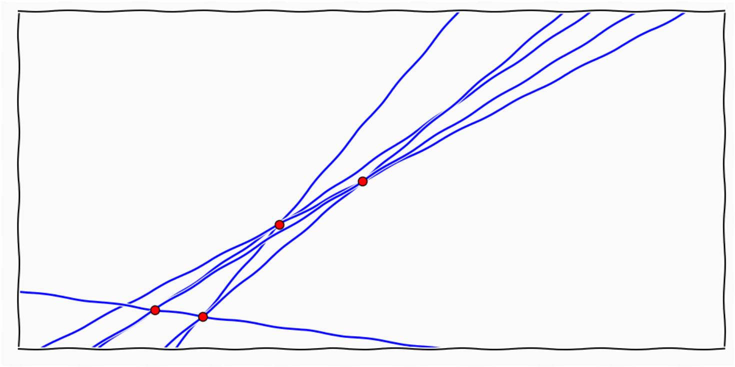
Linear Regression



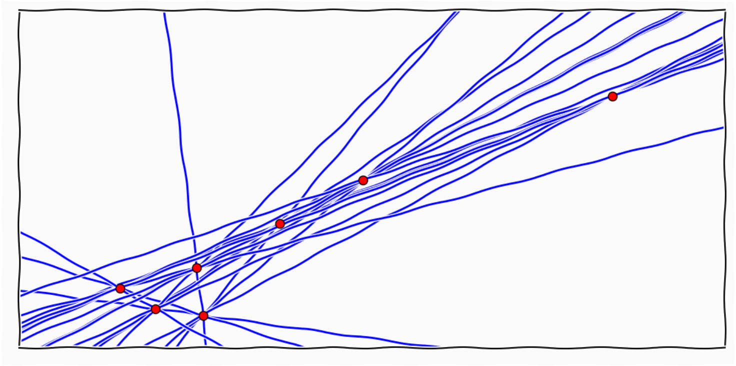
Linear Regression



Linear Regression



Linear Regression



Linear Regression: High School

$$\underbrace{\mathbf{A}}_{m\times n}\underbrace{\mathbf{x}}_{n\times 1} = \underbrace{\mathbf{b}}_{m\times 1}$$

- Over-determined $m > n$

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} \sum_{i}^{m} (b_i - \mathbf{A}_{i:}\mathbf{x})^2$$

Linear Regression: High School

$$\underbrace{\mathbf{A}}_{m\times n}\underbrace{\mathbf{x}}_{n\times 1} = \underbrace{\mathbf{b}}_{m\times 1}$$

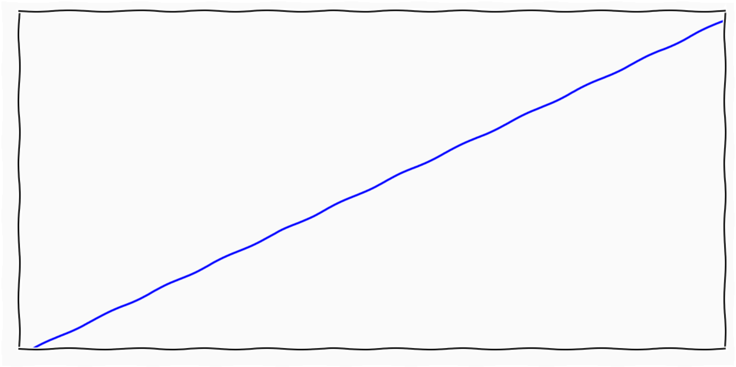
- Over-determined $m > n$

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \sum_{i}^{m} (b_i - \mathbf{A}_{i:}\mathbf{x})^2$$

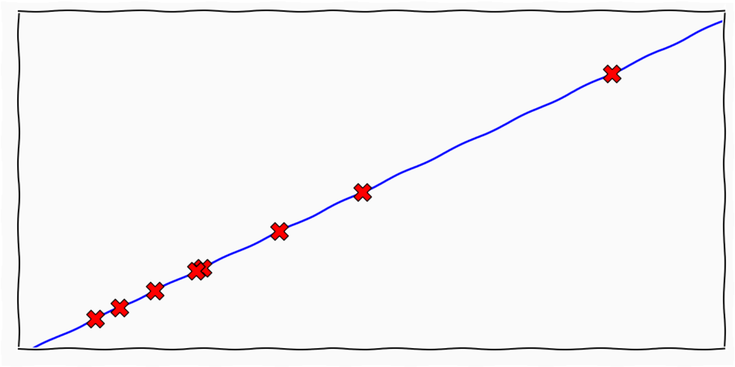
- Under-determined $m < n$

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \sum_{i}^{m} (b_i - \mathbf{A}_{i:}\mathbf{x})^2 + \lambda \mathbf{x}^{\mathrm{T}}\mathbf{x}$$

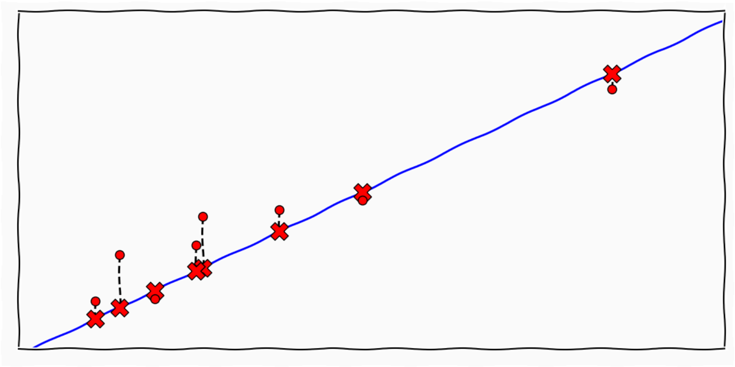
Linear Regression: Modelling



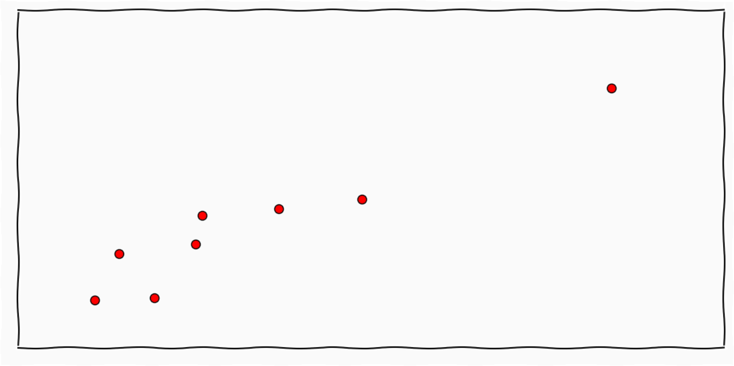
Linear Regression: Modelling



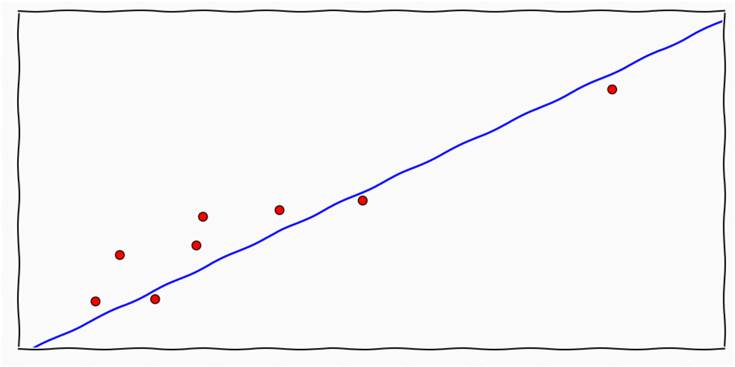
Linear Regression: Modelling



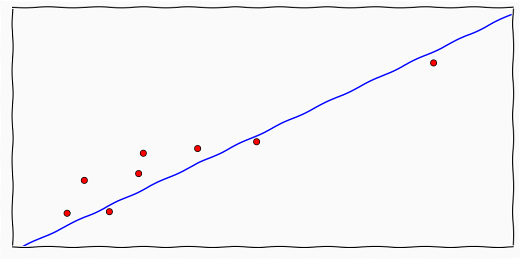
Linear Regression: Modelling



Linear Regression: Modelling



Linear Regression: Modelling



$$y_i = \mathbf{w}^{\mathrm{T}} \mathbf{x}_i + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, \beta^{-1})$$

Belief/Hypothesis

$$y = \mathbf{w}^{\mathrm{T}}\mathbf{x} + \epsilon$$

Belief/Hypothesis

$$y = \mathbf{w}^{\mathrm{T}}\mathbf{x} + \epsilon$$

$$y - \mathbf{w}^{\mathrm{T}}\mathbf{x} = \epsilon$$

Belief/Hypothesis

$$y = \mathbf{w}^{\mathrm{T}}\mathbf{x} + \epsilon$$

$$y - \mathbf{w}^{\mathrm{T}}\mathbf{x} = \epsilon$$

$$y - \mathbf{w}^{\mathrm{T}}\mathbf{x} \sim \mathcal{N}(\epsilon|0, \beta^{-1}I) = \left(\frac{\beta}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}(\epsilon-0)\beta(\epsilon-0)}$$

Belief/Hypothesis

$$y = \mathbf{w}^\mathrm{T}\mathbf{x} + \epsilon$$

$$y - \mathbf{w}^\mathrm{T}\mathbf{x} = \epsilon$$

$$y - \mathbf{w}^\mathrm{T}\mathbf{x} \sim \mathcal{N}(\epsilon|0, \beta^{-1}I) = \left(\frac{\beta}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}(\epsilon-0)\beta(\epsilon-0)}$$

$$\Rightarrow \mathcal{N}(y - \mathbf{w}^\mathrm{T}\mathbf{x}|0, \beta^{-1}I) = \left(\frac{\beta}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}(y-\mathbf{w}^\mathrm{T}\mathbf{x})\beta(y-\mathbf{w}^\mathrm{T}\mathbf{x})}$$

Belief/Hypothesis

$$y = \mathbf{w}^{\mathrm{T}}\mathbf{x} + \epsilon$$

$$y - \mathbf{w}^{\mathrm{T}}\mathbf{x} = \epsilon$$

$$y - \mathbf{w}^{\mathrm{T}}\mathbf{x} \sim \mathcal{N}(\epsilon|0, \beta^{-1}I) = \left(\frac{\beta}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}(\epsilon-0)\beta(\epsilon-0)}$$

$$\Rightarrow \mathcal{N}(y - \mathbf{w}^{\mathrm{T}}\mathbf{x}|0, \beta^{-1}I) = \left(\frac{\beta}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}(y-\mathbf{w}^{\mathrm{T}}\mathbf{x})\beta(y-\mathbf{w}^{\mathrm{T}}\mathbf{x})}$$

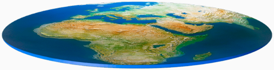
$$\Rightarrow \mathcal{N}(y - \mathbf{w}^{\mathrm{T}}\mathbf{x}|0, \beta^{-1}I) = \mathcal{N}(y|\mathbf{w}^{\mathrm{T}}\mathbf{x}, \beta^{-1}I)$$

Belief/Hypothesis

$$y = \mathbf{w}^{\mathrm{T}}\mathbf{x} + \epsilon$$
$$y - \mathbf{w}^{\mathrm{T}}\mathbf{x} = \epsilon$$
$$y - \mathbf{w}^{\mathrm{T}}\mathbf{x} \sim \mathcal{N}(\epsilon|0, \beta^{-1}I) = \left(\frac{\beta}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}(\epsilon-0)\beta(\epsilon-0)}$$

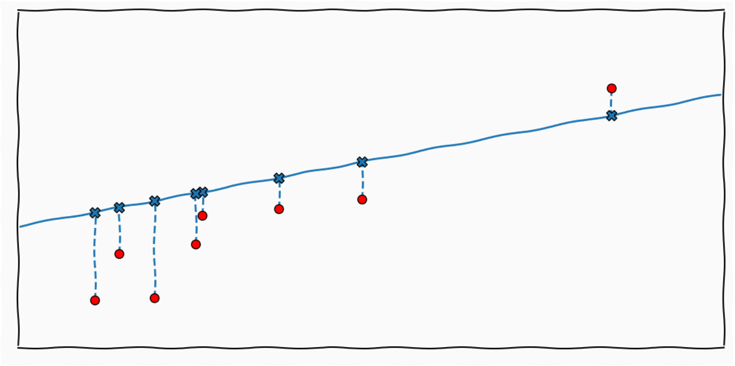
$$\Rightarrow \mathcal{N}(y - \mathbf{w}^{\mathrm{T}}\mathbf{x}|0, \beta^{-1}I) = \left(\frac{\beta}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}(y-\mathbf{w}^{\mathrm{T}}\mathbf{x})\beta(y-\mathbf{w}^{\mathrm{T}}\mathbf{x})}$$
$$\Rightarrow \mathcal{N}(y - \mathbf{w}^{\mathrm{T}}\mathbf{x}|0, \beta^{-1}I) = \mathcal{N}(y|\mathbf{w}^{\mathrm{T}}\mathbf{x}, \beta^{-1}I)$$
$$\Rightarrow p(y|\mathbf{w}, \mathbf{x}) = \mathcal{N}(y|\mathbf{w}^{\mathrm{T}}\mathbf{x}, \beta^{-1}I)$$

Likelihood

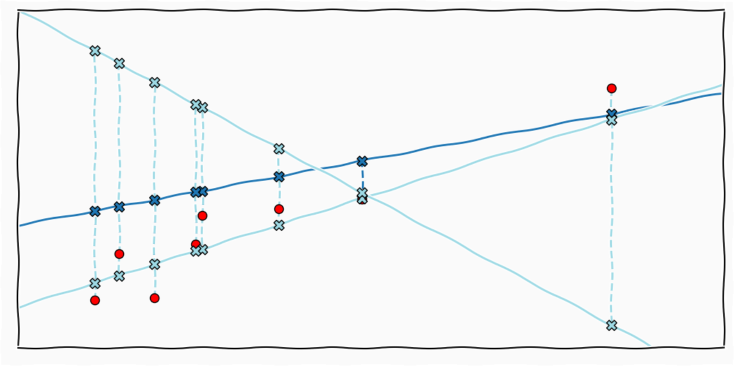


"Given the premise that the earth is flat, how supportive do I believe observations y are of this?"

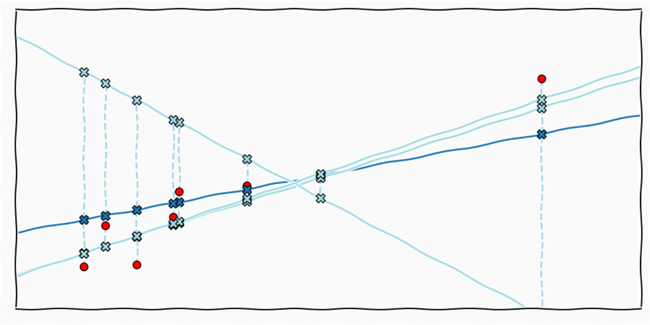
Linear Regression: Modelling



Linear Regression: Modelling

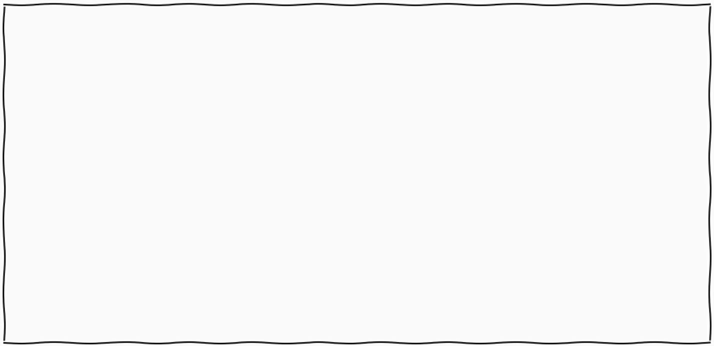


Linear Regression: Likelihood

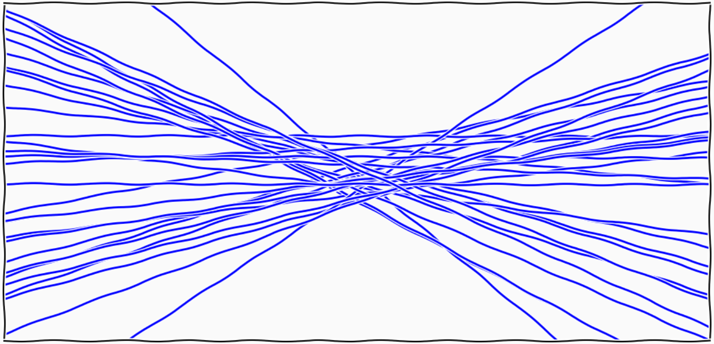


$$p(y|\mathbf{w},\mathbf{x}) = \mathcal{N}(y|\mathbf{w}^{\mathrm{T}}\mathbf{x}, \beta^{-1}I)$$

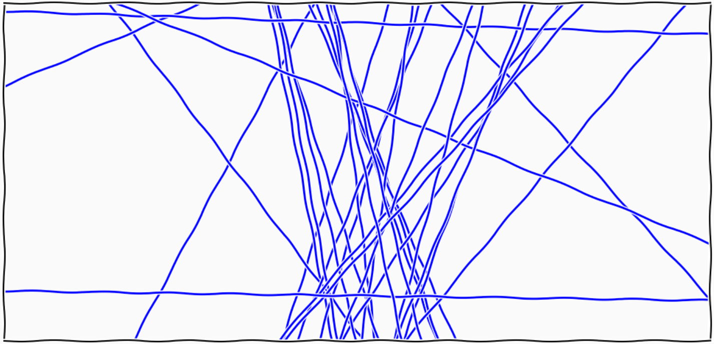
But I don't believe in a flat earth



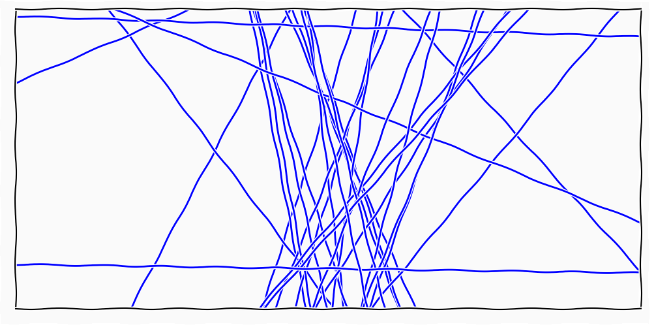
Prior



Prior



Prior



$$w \sim \mathcal{N}(0, 2)$$

Prior

"Well this is how much credibility belief I give to the hypothesis that the earth is flat"

Two opposing theories

That French Dude Again

It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge.
– Laplace Laplace, 1814b

Bayes' "Rule"

$$p(y, w) = p(y|w)p(w)$$

Bayes' "Rule"

$$p(y, w) = p(y|w)p(w)$$
$$p(y, w) = p(w|y)p(y)$$

Bayes' "Rule"

$$
\begin{aligned}
p(y, w) &= p(y|w)p(w) \\
p(y, w) &= p(w|y)p(y) \\
p(w|y)p(y) &= p(y|w)p(w)
\end{aligned}
$$

Bayes' "Rule"

$$
\begin{aligned}
p(y, w) &= p(y|w)p(w) \\
p(y, w) &= p(w|y)p(y) \\
p(w|y)p(y) &= p(y|w)p(w) \\
p(w|y) &= \frac{p(y|w)p(w)}{p(y)}
\end{aligned}
$$

Bayes' "Rule"

$$
\begin{aligned}
p(y, w) &= p(y|w)p(w) \\
p(y, w) &= p(w|y)p(y) \\
p(w|y)p(y) &= p(y|w)p(w) \\
p(w|y) &= \frac{p(y|w)p(w)}{p(y)} \\
&= \frac{p(y|w)p(w)}{\int p(y|w)p(w)\mathrm{d}w}
\end{aligned}
$$

Semantics

$$p(w \mid y) = \frac{p(y \mid w)p(w)}{\int p(y \mid w)p(w)\mathrm{d}w}$$

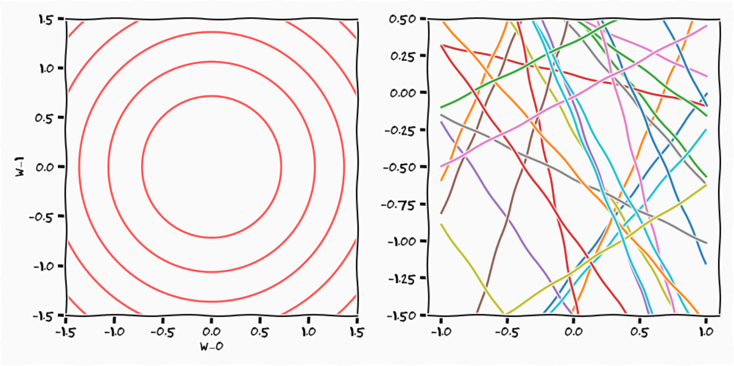
Likelihood How much evidence is there in the data for a specific hypothesis

Prior What are my beliefs about different hypothesis

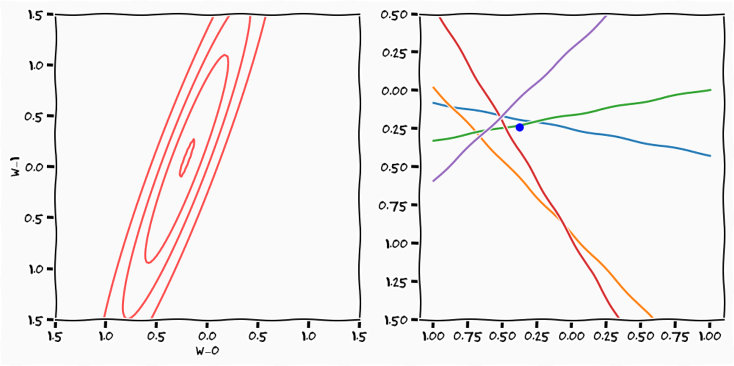
Posterior What is my updated belief after having seen data

Evidence What is my belief about any observations

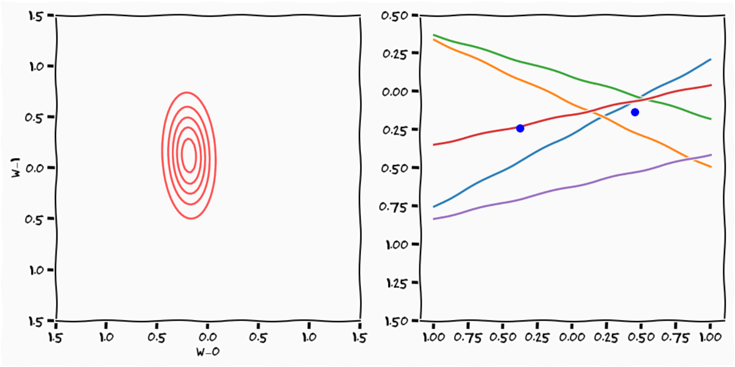
Linear Regression Example

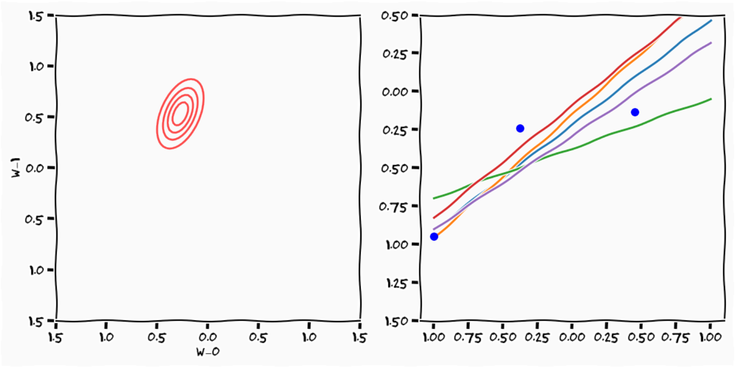


Linear Regression Example

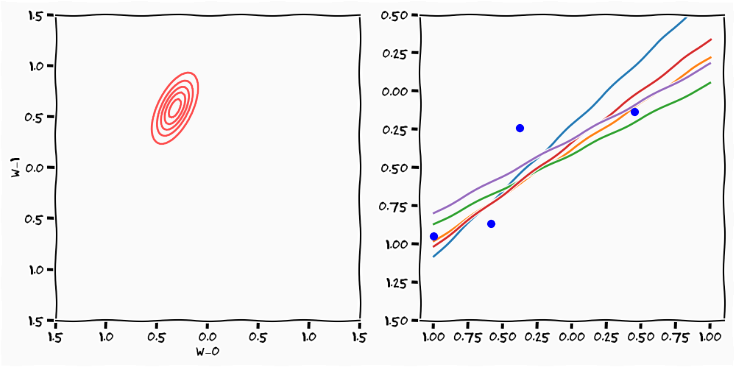


Linear Regression Example

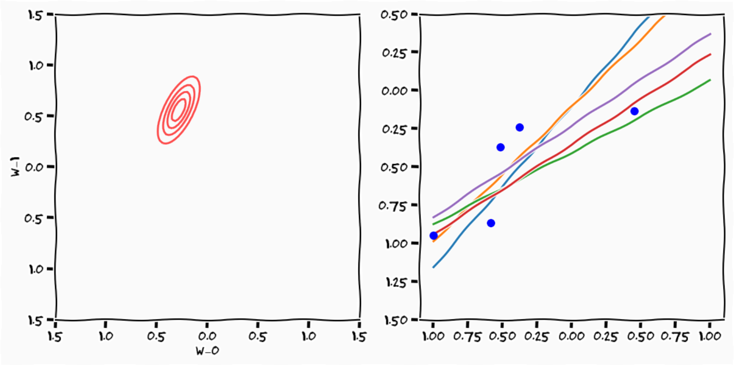



1.5
1.0
0.5
0.0
0.5
1.0
1.5
w_1
1.5 1.0 0.5 0.0 0.5 1.0 1.5
w_0
0.50
0.25
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.00 0.75 0.50 0.25 0.00 0.25 0.50 0.75 1.00

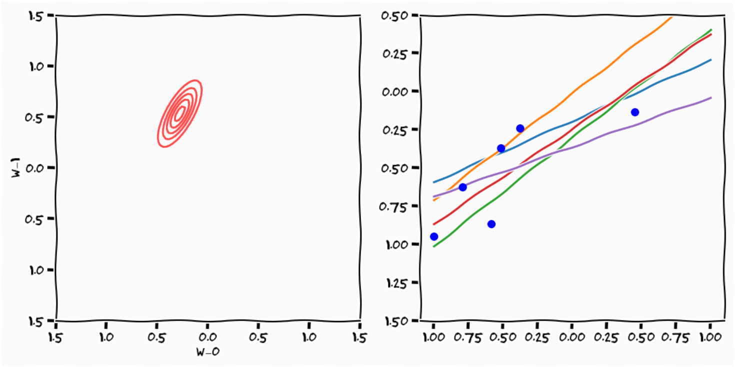
Linear Regression Example



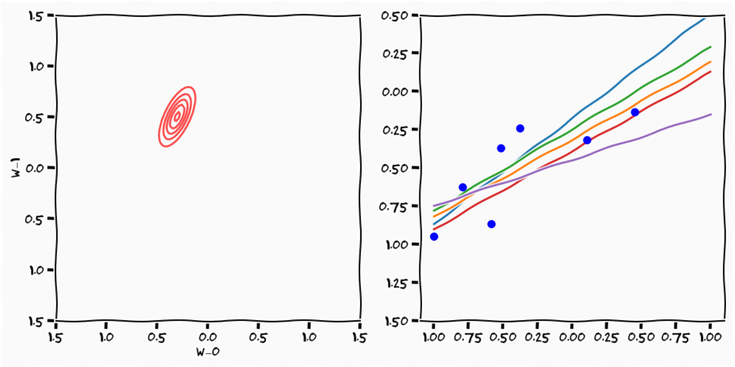
Linear Regression Example



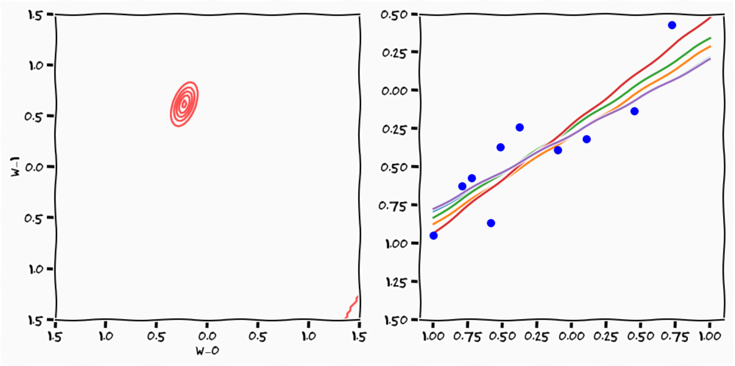
Linear Regression Example



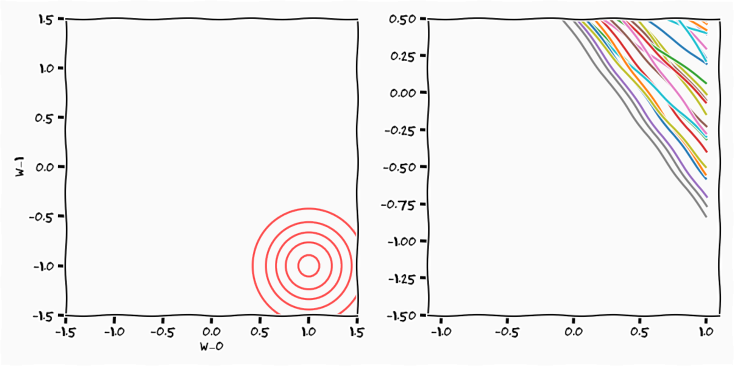
Linear Regression Example



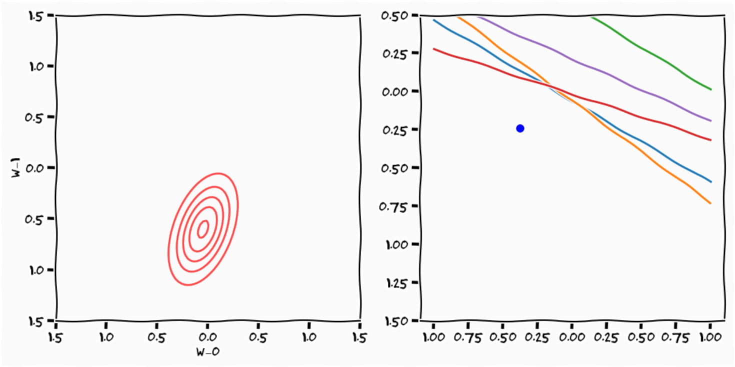
Linear Regression Example



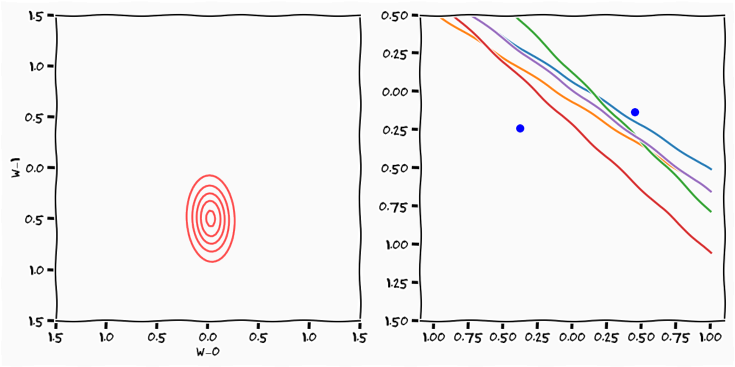
Linear Regression Example



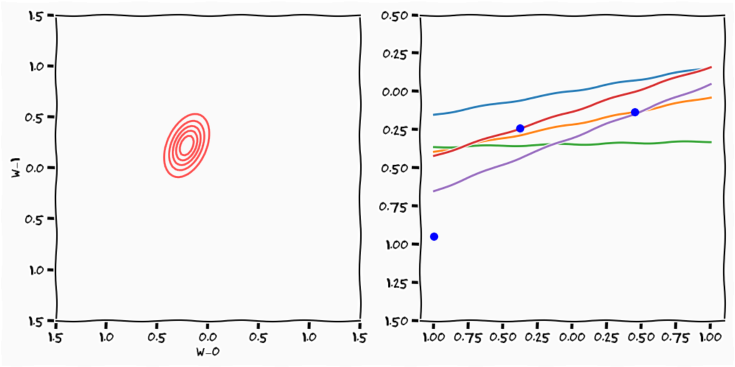
Linear Regression Example



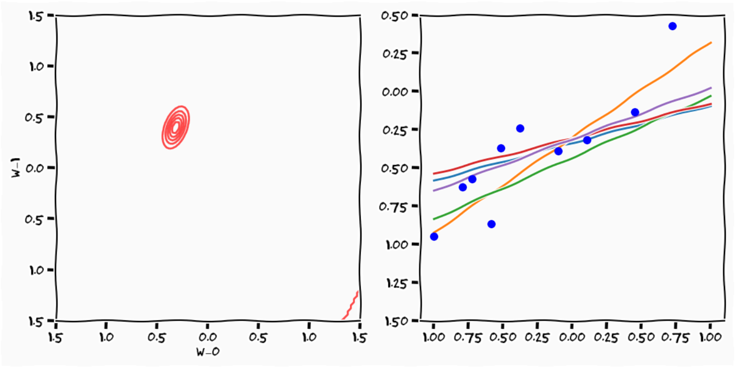
Linear Regression Example



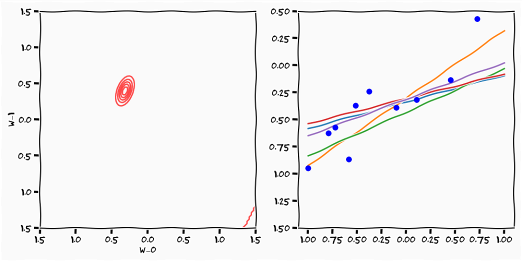
Linear Regression Example

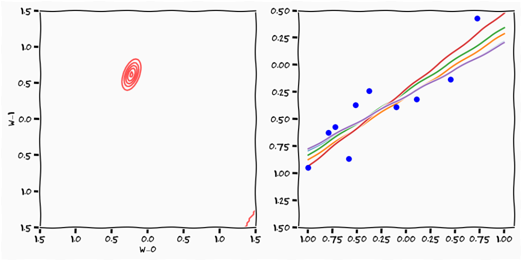


Linear Regression Example

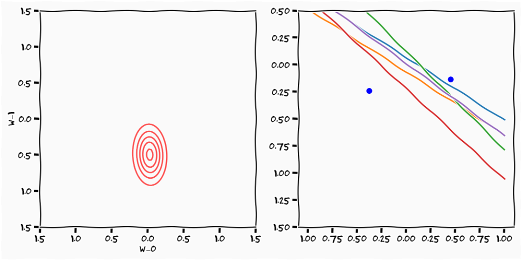


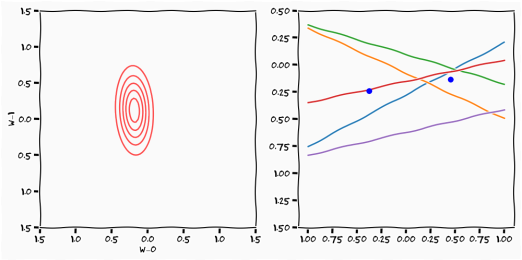
There is overwhelming evidence



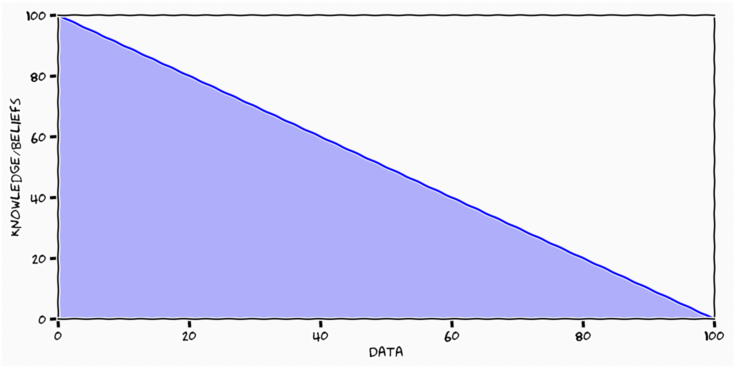


Knowledge is Relative

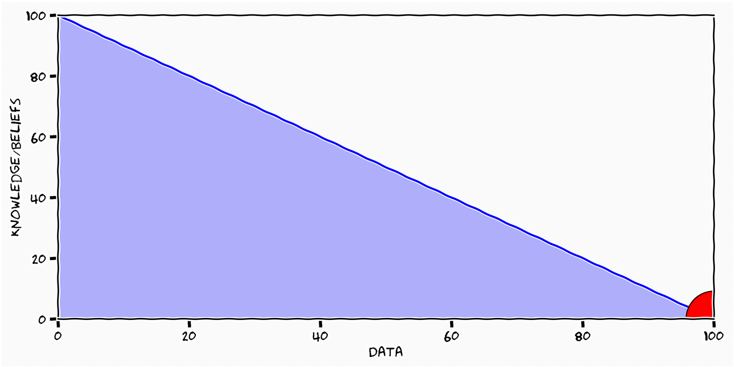




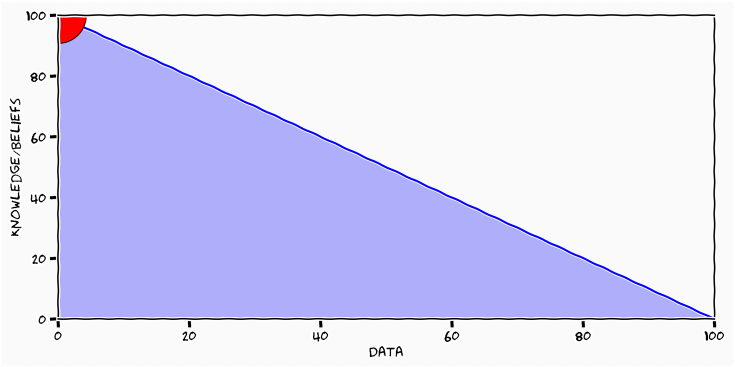
Data and Knowledge



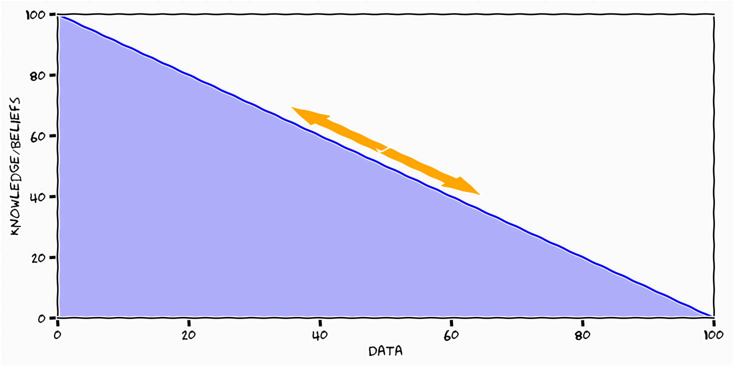
Data and Knowledge



Data and Knowledge

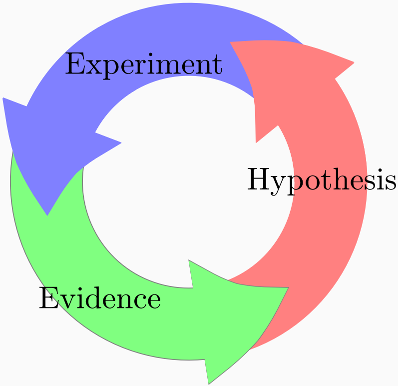


Data and Knowledge



Scientific Principle

Scientific Principle



Where do we start

- Inductive Reasoning

  Observation $\rightarrow$ Pattern $\rightarrow$ Hypothesis $\rightarrow$ Theory

Where do we start

- Inductive Reasoning

  Observation $\rightarrow$ Pattern $\rightarrow$ Hypothesis $\rightarrow$ Theory

- Deductive Reasoning

  Theory $\rightarrow$ Hypothesis $\rightarrow$ Observation $\rightarrow$ Confirmation

Deductive Science

"Science should attempt to disprove a theory, rather than attempt to continually support theoretical hypotheses."

– Karl Popper The Logic of Scientific Discovery

Deductive Science

1. Facilitate viewing implications of Hypothesis in observation space

$$p(w) \rightarrow p(w \mid y)$$

2. Facilitate selection procedure of Hypothesis preference

$$w_1 \succ w_2 \quad p(y \mid w_1) = p(y \mid w_2)$$

What is a good hypothesis?

"In so far as a scientific statement speaks about reality, it must be falsifiable: and in so far as it is not falsifiable, it does not speak about reality."
– Karl Popper The Logic of Scientific Discovery

What is a good hypothesis?

"A theory that explains everything, explains nothing"
– Karl Popper The Logic of Scientific Discovery

Logic vs. Probability

$$P \to \neg Q$$

$$\neg P$$

$$Q$$

$$p(y) = \int p(y|\theta)p(\theta)\mathrm{d}\theta$$

Falsifiability and Occams' Razor

- A hypothesis should be judged based on how easily it can be falsified

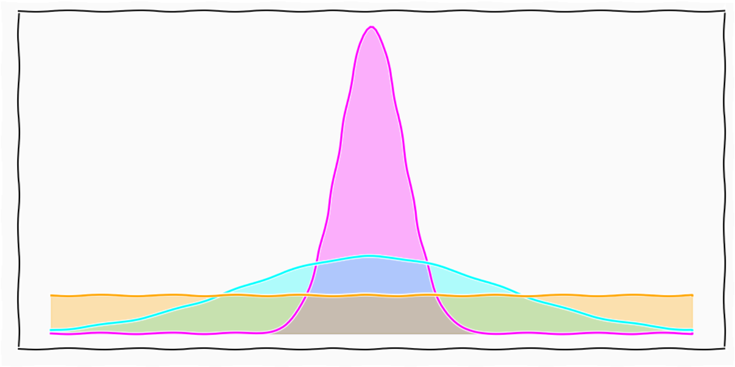
Falsifiability and Occams' Razor

- A hypothesis should be judged based on how easily it can be falsified
- The more general a theory is the more cases/possibilities it allows for falsification

Falsifiability and Occams' Razor

- A hypothesis should be judged based on how easily it can be falsified
- The more general a theory is the more cases/possibilities it allows for falsification
- "The more strongly our framework can differentiate different hypothesis the better it is for falsification"

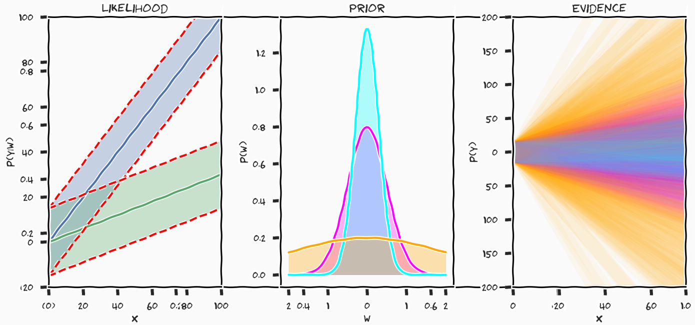
Distributions



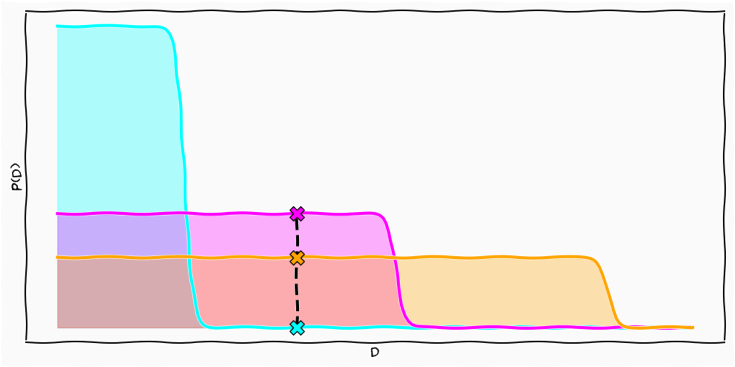
Model Evidence and Occams' Razor

$$p(y) = \int p(y \mid w) p(w) \mathrm{d}w$$

What is can be falsified?



The MacKay Plot Mackay, 1991



Is Machine Learning a Science?

- How to build mathematical models of hypothesis

$$\text{hypothesis} \approx p(w)$$

Is Machine Learning a Science?

- How to build mathematical models of hypothesis

$$\text{hypothesis} \approx p(w)$$

- How to mathematically update our knowledge with data

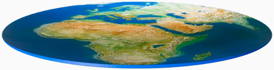
$$p(w \mid y) \approx \frac{p(y \mid w)p(w)}{\int p(y \mid w)p(w)\mathrm{d}w}$$

Summary

Summary

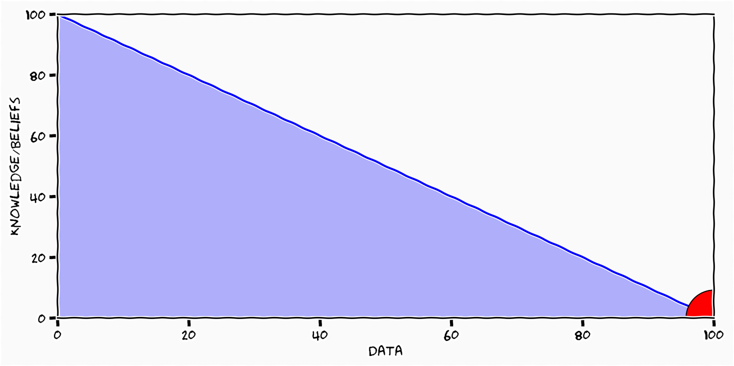
Machine Learning is a framework for combining knowledge with data to recover an interpretation of the data in light of the knowledge.

Where does Knowledge Come From?



Machine Learning and Science

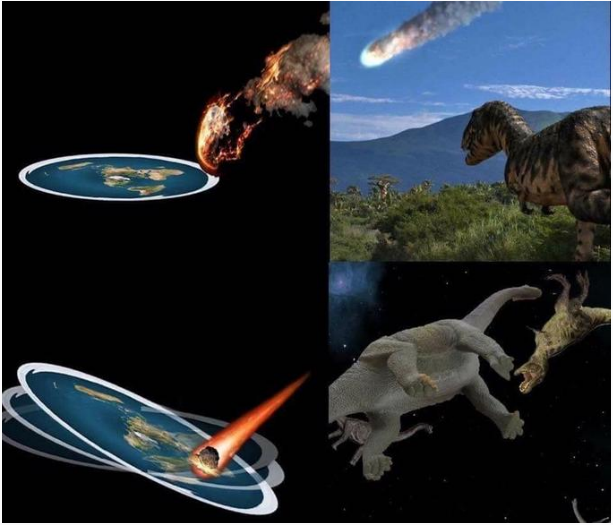
Don't believe the hype



Summary

"There is no logical paths leading to "these laws" they can only be reached by intuition based on something like and intellectual love of the objects of experience"
– Albert Einstein

Thats why they disappeared



eof

References

- Laplace, Pierre Simon (1814a). *A philosophical essay on probabilities*.
- — (1814b). *Théorie analytique des probabilités*. Mme. Ve. Sourcier.
- Mackay, David J C (Dec. 1991). "Bayesian methods for adaptive models ". PhD thesis. California Institute of Technology: California Institute of Technology.
- Popper, K.R. (1959). *The Logic of Scientific Discovery*. ISSR library. Routledge.