

# ML Week

## Introduction

Jeff Abrahamson

23–25 novembre 2016

# About Machine Learning

**The literature is overwhelmingly in English**

**Disclaimer: Time is short**

## Definition

# Supervised

# Unsupervised

# Reinforcement



# Curse of Dimensionality

**Machine learning is not magic**

# Machine learning is mathematics

## Mostly, it's these maths:

- Probability
- Statistics
- Linear algebra
- Optimisation theory
- Differential calculus

# Probability

## Addition rule: independent events

$$\Pr(A \cup B) = \Pr(A) + \Pr(B)$$

## Addition rule: dependent events

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

## Multiplication rule: independent events

$$\Pr(A \cap B) = \Pr(A) \Pr(B)$$



## Multiplication rule: dependent events

$$\Pr(A \cap B) = \Pr(A \mid B) \Pr(B)$$

# Conditional probability

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

# Conditional probability

$$\cup_j A_j = A \quad \wedge \quad A_i \cap A_j = \emptyset \implies$$

$$P(A_1 | B) = \frac{\Pr(B | A_1) \Pr(A_1)}{\sum_i \Pr(B | A_1) \Pr(A_1) + \cdots + \Pr(B | A_k) \Pr(A_k)}$$

# Statistics

# What is Statistics

- 1 Identify a question or problem.
- 2 Collect relevant data on the topic.
- 3 Analyze the data.
- 4 Form a conclusion.

# What is Statistics

- 1 Identify a question or problem.
- 2 Collect relevant data on the topic.
- 3 Analyze the data.
- 4 Form a conclusion.

Sadly, sometimes people forget 1.

# What is Statistics

- 1 Identify a question or problem.
- 2 Collect relevant data on the topic.
- 3 Analyze the data.
- 4 Form a conclusion.

Statistics is about making 2–4 efficient, rigorous, and meaningful.

*OpenIntro Statistics, 2nd edition, D. Diez, C. Barr, M. Çetinkaya-Rundel, 2013.*

# What is data science?

(Exercise: Is this the same question as the last slide?)

- 1 Define the question of interest
- 2 Get the data
- 3 Clean the data
- 4 Explore the data
- 5 Fit statistical models
- 6 Communicate the results
- 7 Make your analysis reproducible



# What is data science?

(Exercise: Is this the same question as the last slide?)

- 1 Define the question of interest
- 2 Get the data
- 3 Clean the data
- 4 Explore the data
- 5 Fit statistical models
- 6 Communicate the results
- 7 Make your analysis reproducible

What the public thinks.

# What is data science?

(Exercise: Is this the same question as the last slide?)

- 1 Define the question of interest
- 2 Get the data
- 3 Clean the data
- 4 Explore the data
- 5 Fit statistical models
- 6 Communicate the results
- 7 Make your analysis reproducible

Where we spend most of our time.

# What is data science?

(Exercise: Is this the same question as the last slide?)

- 1 Define the question of interest
- 2 Get the data
- 3 Clean the data
- 4 Explore the data
- 5 Fit statistical models
- 6 Communicate the results
- 7 Make your analysis reproducible

The easiest part to forget.

# What is data science?

*[http://simplystatistics.org/2015/03/17/  
data-science-done-well-looks-easy-and-that-is-a-big-  
problem-for-data-scientists/](http://simplystatistics.org/2015/03/17/data-science-done-well-looks-easy-and-that-is-a-big-problem-for-data-scientists/)*

# What is data science?



# Anecdote

Some properties of anecdote:

- is data
- haphazardly collected
- is generally not representative
- sometimes result of selective retention
- does not accumulate to be representative
- might be true (by chance)
- is ok to use as hypothesis, but be clear that hypothesis is anecdote

# Study Types

- Observational
- Experimental

# Study Types

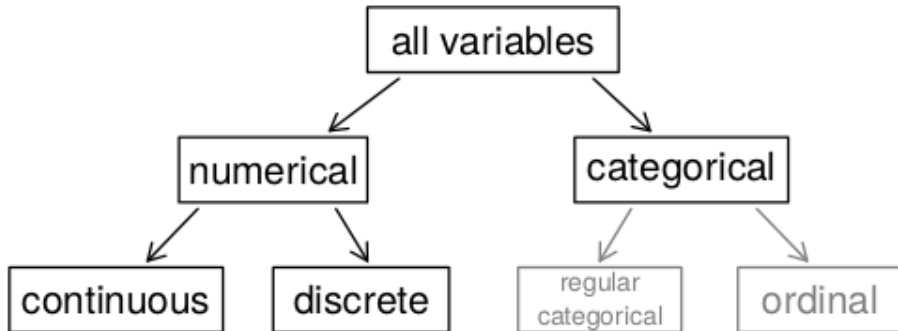
- Observational
- Experimental

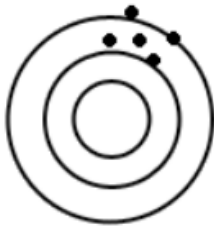
What can go wrong?

- Forgetting that association  $\neq$  causation
- Not random
- Confounding variables

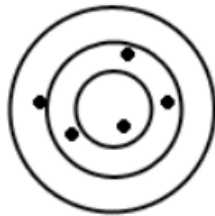


# Variable types





High bias, low variance



Low bias, high variance



High bias, high variance



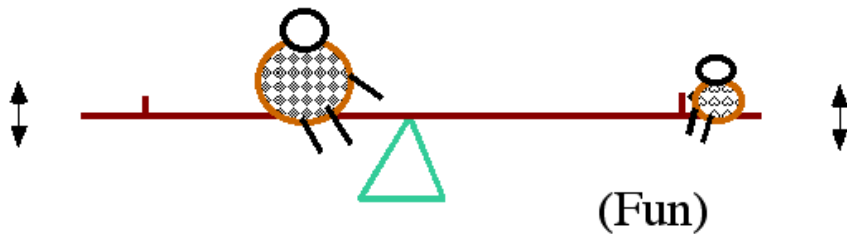
Low bias, low variance

# Mean

- Weighted and unweighted
- Centroid to physicists

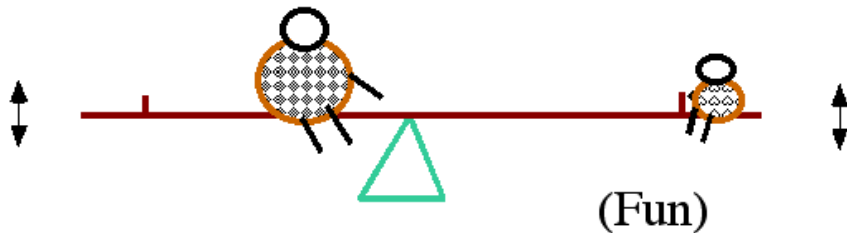
# Mean

- Weighted and unweighted
- Centroid to physicists



# Mean

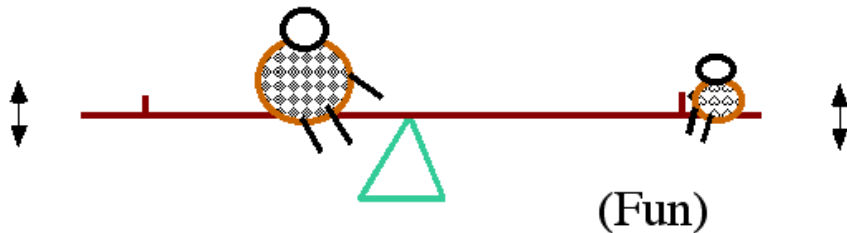
- Weighted and unweighted
- Centroid to physicists



$$\mu = E(X) = \sum w_i x_i = \mathbf{w} \cdot \mathbf{x}$$

# Mean

- Weighted and unweighted
- Centroid to physicists



$$\mu = E(X) = \sum \Pr(X = x_i)x_i$$

# Mean

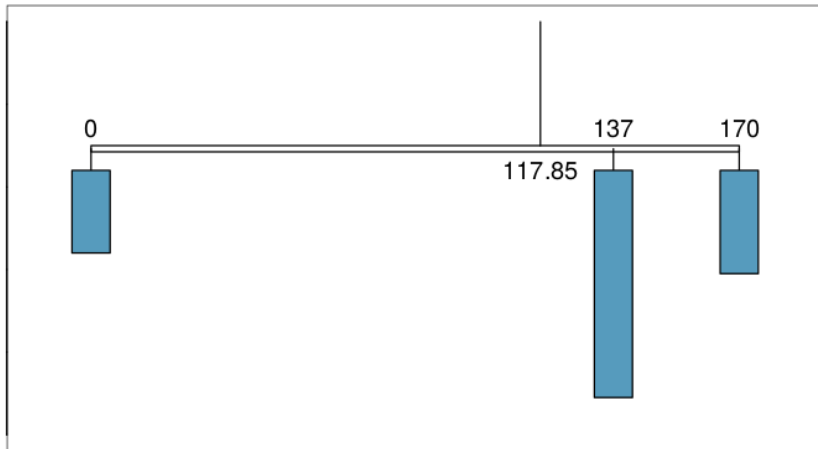
- Weighted and unweighted
- Centroid to physicists

$$\mu = E(X) = \int xf(x) dx$$

*<http://telescopes.stardate.org/images/research/teeter-totter/TT4.gif>*

# Mean

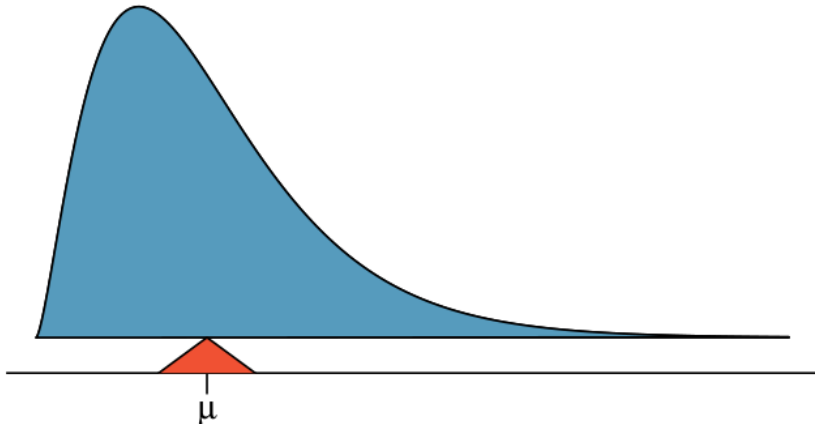
- Weighted and unweighted
- Centroid to physicists





# Mean

- Weighted and unweighted
- Centroid to physicists



**Deviation** is distance from mean.

**Variance** is mean square of deviations

**Standard deviation** is square root of variance

$$s^2 = \frac{(\bar{x} - x_1)^2 + \cdots (\bar{x} - x_n)^2}{n - 1}$$

# Population statistics

$$\sigma^2 = \frac{(\bar{X} - x_1)^2 + \cdots (\bar{X} - x_n)^2}{n}$$

$$\text{Var}(X) = \sigma^2 = (\bar{x} - x_1)^2 \Pr(X = x_1) + \cdots (\bar{x} - x_n)^2 \Pr(X = x_n)$$

# Words

Given a distribution  $X$ , the

**probability distribution function (pdf)** (continuous) or **probability mass function (pmf)** is the probability that the variate has value  $x$ :

$$\Pr(a \leq X \leq b)$$

or

$$\Pr(X = a)$$



Given a distribution  $X$ , the

**cumulative probability function (cdf)** is the probability that the variate is less than  $x$ :

$$\Pr(X \leq x) = \int_{-\infty}^x \text{pdf}(x) \, dx \text{ or } \sum_{i \leq x} \Pr(X = x)$$

# Words

Given a distribution  $X$ , the

**percent point function (ppf)** is the inverse of the cdf. Given a probability, what's  $x$ ? Also called the **inverse distribution**.

Given a distribution  $X$ , the

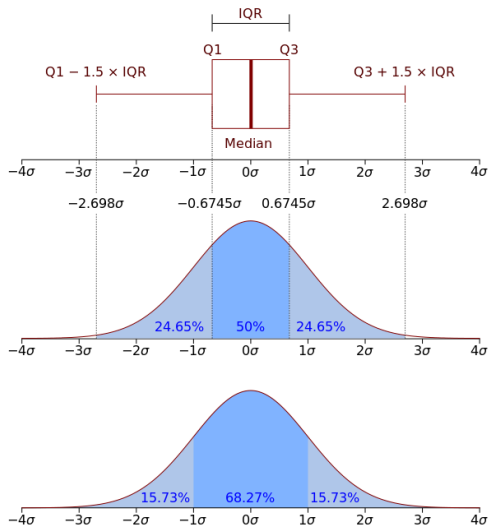
**survival function (sf)** is the probability that the variate takes a value greater than  $x$ :

$$ss(x) = \Pr(X > x) = 1 - \text{cdf}(x)$$

Given a distribution  $X$ , the

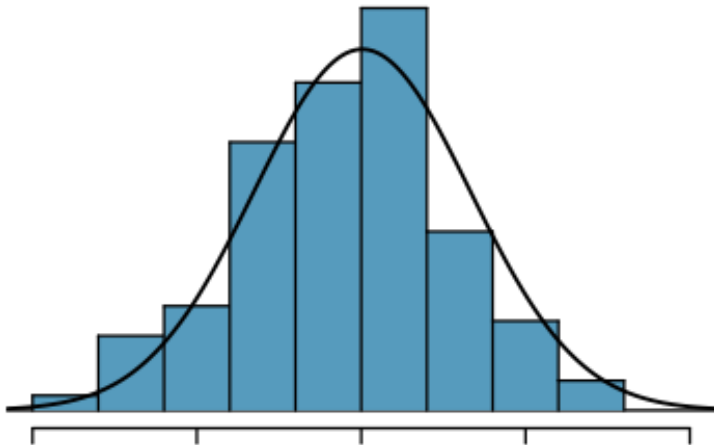
**inverse survival function (isf)** is the inverse of the survival function:

$$\text{isf}(\alpha) = \text{ppf}(1 - \alpha)$$



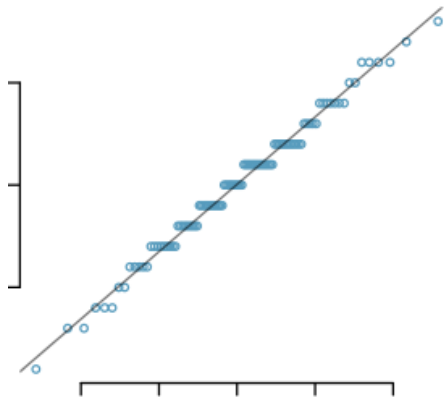
# Evaluating Normal Approximations

Easy technique 1: visually compare to normal plot.



# Evaluating Normal Approximations

Easy technique 2: normal probability plot.



Also known as a quantile-quantile plot.

$$\bar{x} \neq \mu$$

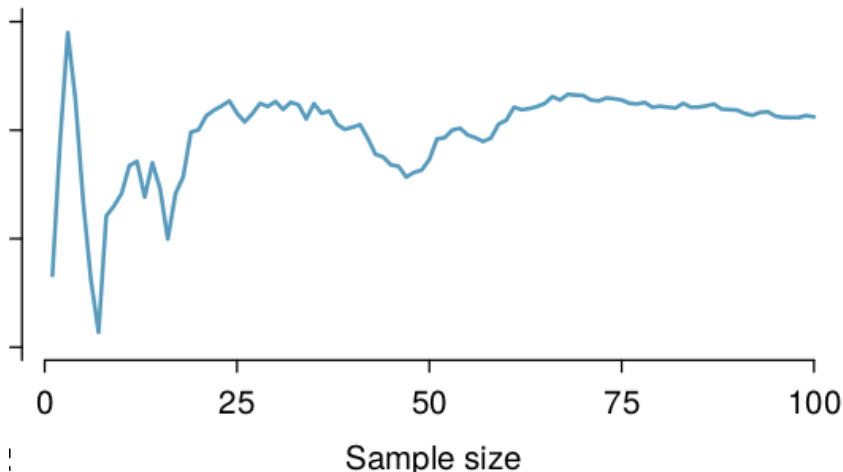


# Inference Concepts

**Running mean.** Sequence of partial sums (divided by number in sum).

# Inference Concepts

**Running mean.** Sequence of partial sums (divided by number in sum).



# Inference Concepts

**Running mean.** Sequence of partial sums (divided by number in sum).

**Sampling variation.** Change of  $\bar{x}$  from one sample to the next.

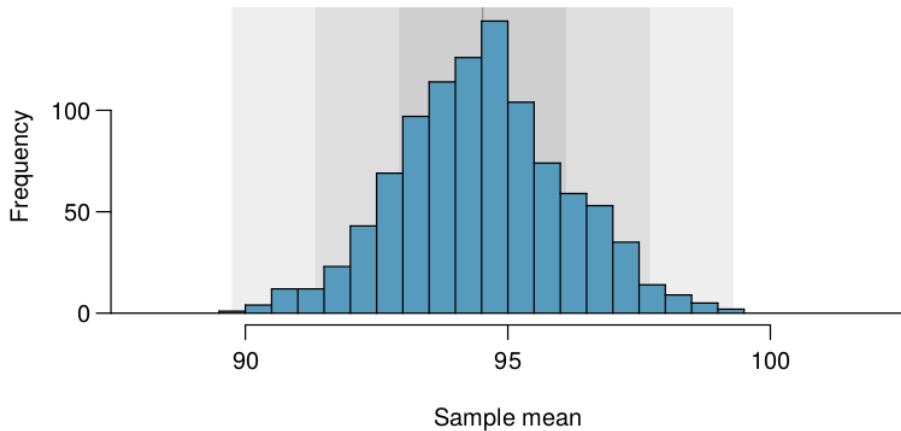
# Inference Concepts

**Running mean.** Sequence of partial sums (divided by number in sum).

**Sampling variation.** Change of  $\bar{x}$  from one sample to the next.

**Sampling distribution.** The distribution of possible point samples of a fixed size from a given population.

# Sampling distribution



# Confidence intervals

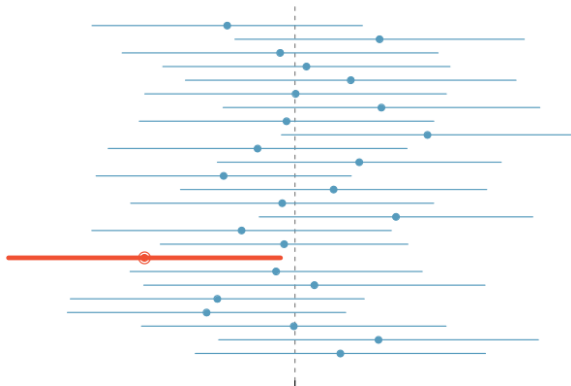
Sample  $n$  points, choose an interval around the sample mean.

A 95% confidence interval means if we sample repeatedly, about 95% of the samples will contain the population mean.

# Confidence intervals

Sample  $n$  points, choose an interval around the sample mean.

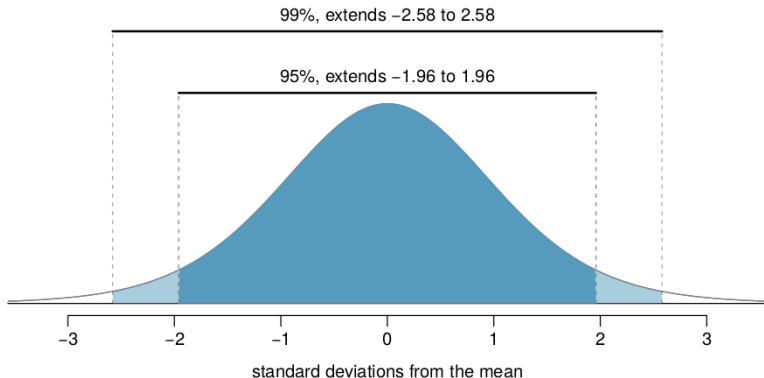
A 95% confidence interval means if we sample repeatedly, about 95% of the samples will contain the population mean.



# Confidence intervals

Sample  $n$  points, choose an interval around the sample mean.

A 95% confidence interval means if we sample repeatedly, about 95% of the samples will contain the population mean.





# Linear Algebra

# Linear algebra: basics

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$$

# Linear algebra: basics

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix}$$
$$= \left\{ \begin{matrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{matrix} \right\} \in \mathbb{R}^{n \times n}$$

# Linear algebra: basics

$$u + v = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{pmatrix}$$

# Linear algebra: basics

$$\alpha \mathbf{V} = \begin{pmatrix} \alpha v_1 \\ \alpha v_2 \\ \vdots \\ \alpha v_n \end{pmatrix} \quad (\alpha \in \mathbb{R})$$

# Linear algebra: basics

$$\| \mathbf{v} \| = \sqrt{v_1^2 + \cdots + v_n^2}$$

# Linear algebra: basics

$$u \cdot v = u_1 \cdot v_1 + \cdots + u_n \cdot v_n$$

$$= \|u\| \|v\| \cos \theta$$

# Linear algebra: basics

$$C = A + B \iff c_{ij} = a_{ij} + b_{ij}$$

$$C = AB \iff c_{ij} = \sum_k a_{ik} b_{kj}$$

$$A = B^T \iff a_{ij} = b_{ji}$$

$$AA^{-1} = A^{-1}A = \text{diag}(1)$$



# Linear algebra: transformations

$$Ax = y \qquad f = T_A : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

$$x = A^{-1}Ax = A^{-1}y \qquad f^{-1} = T_{A^{-1}} : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

# Linear algebra: transformations

$B$  is a basis for  $V$  iff any of these conditions are met:

- $B$  is a minimal generating set of  $V$
- $B$  is a maximal set of linearly independent vectors
- Every vector  $v \in V$  can be expressed in a unique way as a sum of  $b_i \in B$

(The conditions are equivalent.)

# Linear algebra: transformations

$B$  is a basis for  $V$  iff any of these conditions are met:

- $B$  is a minimal generating set of  $V$
- $B$  is a maximal set of linearly independent vectors
- Every vector  $v \in V$  can be expressed in a unique way as a sum of  $b_i \in B$

(The conditions are equivalent.)

Bases are not unique.

# Linear algebra: transformations

Eigenvectors, eigenvalues:

$$Av = \lambda v$$

# Linear algebra: transformations

Eigenvectors, eigenvalues:

$$Av = \lambda v$$

$$Av = \lambda 1 v \iff (A - \lambda 1)v = 0$$

# Linear algebra: transformations

Eigenvectors, eigenvalues:

$$Av = \lambda v$$

Some matrices are diagonalisable. Then

$$A = Q\Lambda Q^{-1} \quad \text{with } \Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$$

$$\text{and } Q = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}$$

# Questions?

[ml-week.com/1](http://ml-week.com/1)