Machine Learning

Images and Dimensionality Reduction

Jeff Abrahamson

Cours sur l'année, 2017–2018

Review

What is Machine Learning?

Learning is what we do when we can't explain how.

?

What is Machine Learning?

Learning is what we do when we can't explain how.

- Supervised
- Unsupervised
- Reinforcement

Lots of maths

We'll try to ignore it, but it's there...

- Vector spaces and linear algebra
- Probability
- Statistics
- Optimisation theory
- Differential calculus

The curse of dimensionality.

Data Science

?

Data Science

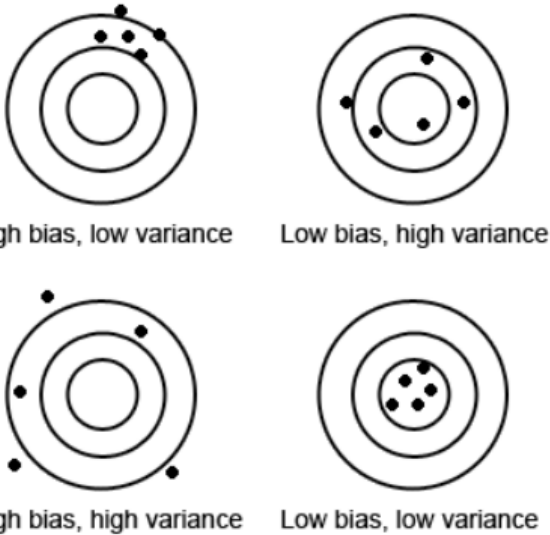
1. Define the question of interest
2. Get the data
3. Clean the data
4. Explore the data
5. Fit statistical models
6. Communicate the results
7. Make your analysis reproducible

Observational vs experimental

Data

Anecdote: it doesn't accumulate to be data.

Data



Features

Feature Engineering

Data

One of K = one-hot encoding

Outliers: don't ignore them!

Feature Engineering

1. Brainstorm
2. Pick some
3. Make them
4. Evaluate
5. Repeat

Easy Features

Text

bag of words

Easy Features

Images

corners, edges, point matching

Linear Regression

?

Linear Regression

Problem: $\{(x_i, y_i)\}$.

Given x, predict $\hat{y}$.

Here y is continuous.

Linear Regression

x: **explanatory** or **predictor** variable.

y: **response** variable.

For some reason, we believe a linear model is a good idea.

Residuals

What's left over.

?

Residuals

What's left over.

$$\text{data} = \text{fit} + \text{residual}$$

Residuals

What's left over.

$$y_i = \hat{y}_i + e_i$$

Residuals

What's left over. Goal: small residuals.

$$\sum e_i^2$$

Logistic regression

?

Logistic regression

- Binary output
- Classification

Logistic regression

- Have: continuous and discrete inputs
- Want: class (0 or 1)

Logistic regression

Logistic (sigmoid, logit) function

$$g(z) = \frac{1}{1 + e^{-z}}$$

One vs Rest, One vs One

One vs Rest, One vs One

- OvR (OvA): compute k classifiers
- OvO: compute $k(k-1)/2$ classifiers

The classifiers give scores, not just in/out answers.

One vs Rest, One vs One

One vs Rest:

Accept the judgement of the classifier with the highest score.

One vs Rest, One vs One

One vs One:

Classifiers vote. Accept the class that gets the most votes.

One vs Rest, One vs One

Advantage: Reduces multi-class classification to single-class classification.

Disadvantage: Classifier scores aren't necessarily comparable. For example, classes may have very different numbers of members.

Hyperparameters

?

Hyperparameters

- The word hyperparameter is not well-defined.
- In most contexts, it is the parameters of the underlying distribution
- In training, we learn the parameters of the model
- We choose the hyperparameters to govern the training
- So we may want to experiment to learn the distribution parameters that best optimise our learned model's performance

Testing

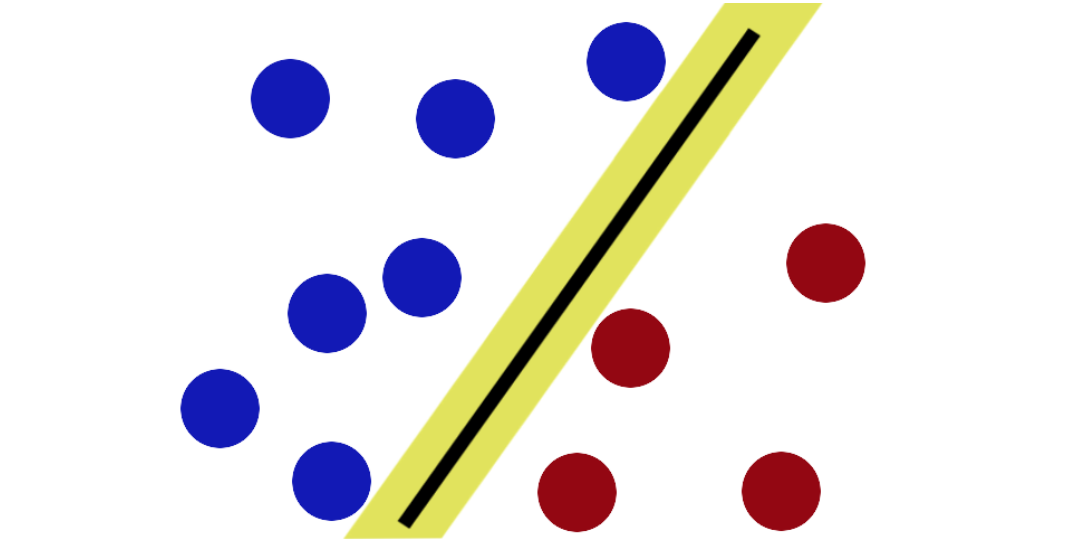
- Set aside (partition) data for testing (e.g., 70% / 30%)
- Learn on training set, test on testing set
- When searching hyperparameters, set aside again (e.g., 60% / 20% / 20%)

SVM

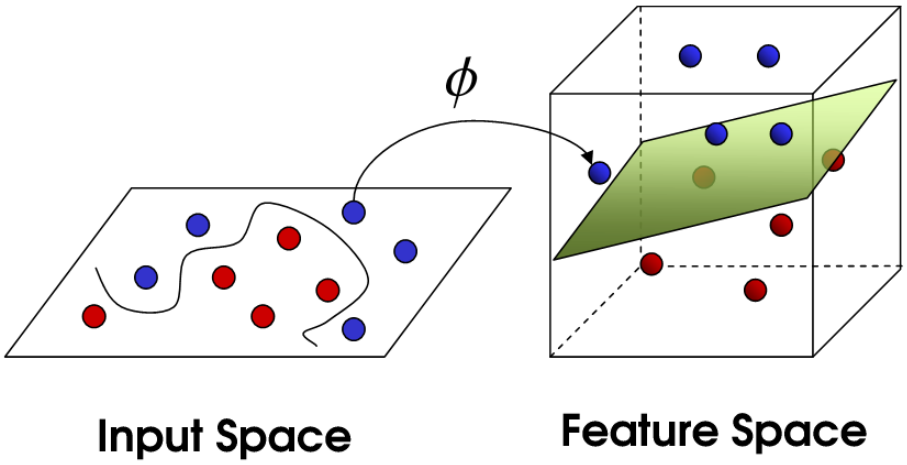
?

copperking@reddit

SVM



SVM



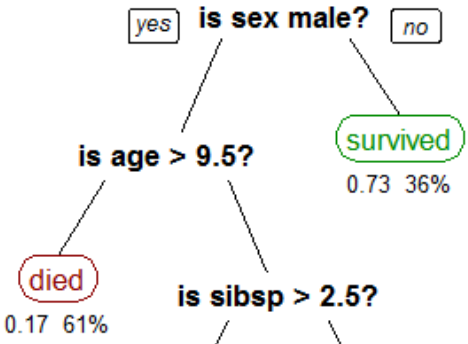
CART

?

Random Forest

?

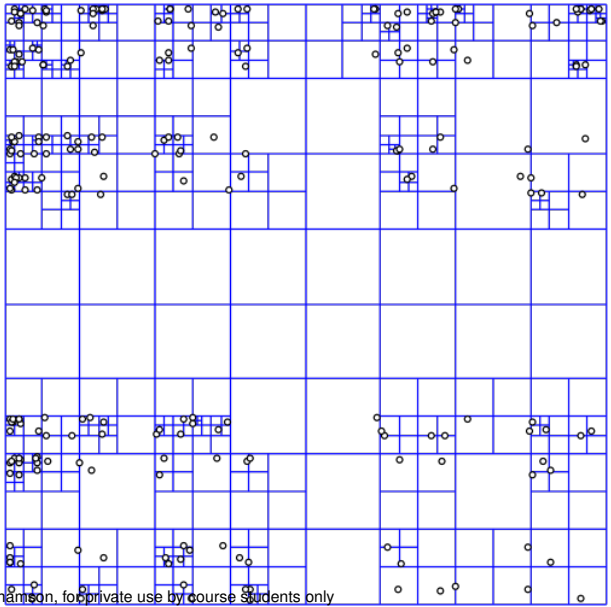
Decision Trees



E.g., passengers died with probability .17 which is 61% of observations

Stephen Milborrow

Quadtree



Decision Trees

Variations

- Classification tree
- Regression tree

CART = classification and regression trees

Decision Trees

Ensemble methods

- Bagging
- Random forest
- Boosted trees *(gradient boosted trees)*
- Rotation forest

Boostrap aggregating = bagging

Bagging

- Increase stability
- Increase accuracy
- Reduce variance
- Avoid overfitting

A type of model averaging (ensemble method).

Boostrap aggregating = bagging

- Training set D of size n
- Sample D *with replacement* to create $D_1, \ldots, D_k$ of size n'
- If $n = n'$, expect $1 - 1/e \approx 63.2\%$ repeats

- Train k models
- Average (regression) or vote (classification)

Random subspace method

attribute bagging = feature bagging

Random subspace method

Bagging (bootstrap aggregation) = resampling to create more data sets, train models on different samples

Attribute bagging = project to create more data sets, train models on different samples

Random forests

Combine bagging with random subspace method

Images

Signal processing

in 2 or 3 dimensions

Images

Details that can matter:

- Illumination
- White balance
- Resolution
- Camera settings (e.g., depth of field)
- Sensor noise
- Compression technology

Images

Challenges:

- Segmentation
- Area of interest detection
- Perspective shifting

Images

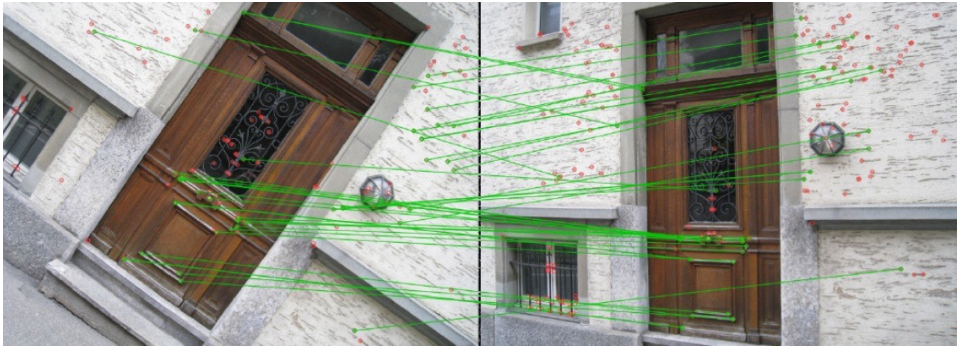
Applications:

- Agriculture: fruit ripening, automated harvesting
- Security: detecting specific people
- Security: detecting accidents (e.g., falls)
- Art: counterfeit detection
- Medicine: assisted surgery
- Image search

Images

Image search (at first):

- Texture
- Colour
- Shape, simple objects



Eddie Bell @ Lyst

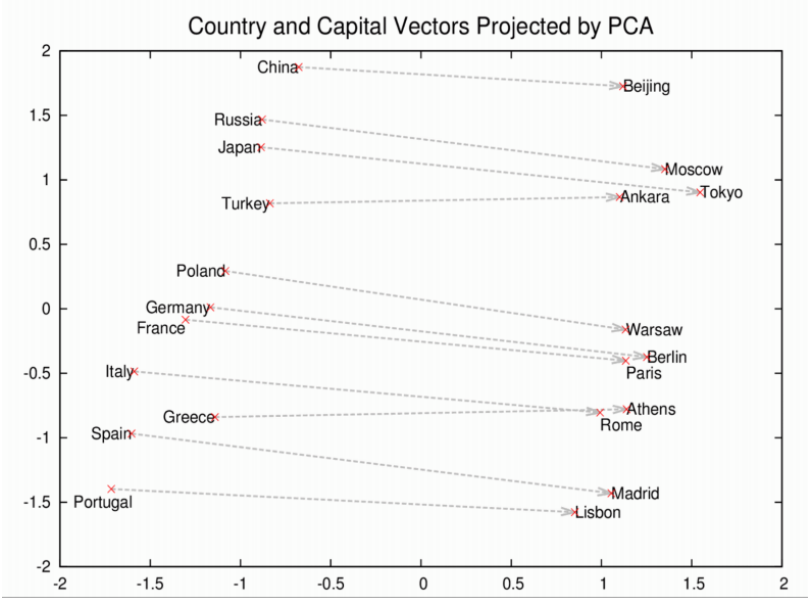
Eddie Bell @ Lyst

Eddie Bell @ Lyst

Eddie Bell @ Lyst

-1	0	+1
-2	0	+2
-1	0	+1

-1	-2	-1
0	0	0
+1	+2	+1



	Term	Similarity
	"shift"	0.933104
	"gown"	0.887743
	"skirt"	0.881672
	"bandage"	0.880162
	"midi"	0.869786

Similar to ‘dress’

Eddie Bell @ Lyst

a group of young girls standing next to each other on the beach

A clock tower with a clock on top of it

Google?

A bunch of bananas hanging from a tree

Google?

PCA

Principle component analysis

Analyse en composantes principales

Motivation

Remember the Curse of Dimensionality?

Principle

- Linear transformations have axes
- Find them (eigenvectors of the covariance matrix)
- Pick the biggest ones

Principle

- Linear transformations have axes
- Find them (eigenvectors of the covariance matrix)
- Pick the biggest ones

Fitting an n-dimensional ellipsoid to the data

Uses

- Exploratory data analysis
- Compression

Also known as

- Discrete Kosambi-Karhunen–Loève transform (KLT) (signal processing)
- Hotelling transform (multivariate quality control)
- Proper orthogonal decomposition (POD) (ME)
- Singular value decomposition (SVD), Eigenvalue decomposition (EVD) (linear algebra)
- Etc.

History

- Invented by Karl Pearson in 1901
- Invented (again) and named by Harold Hotelling in 1930's
- Also known as...

Also known as

- It's a long list, every field uses a different name...