

Machine Learning

Images and Dimensionality Reduction

Jeff Abrahamson

octobre 2020

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

Data

Observational vs experimental

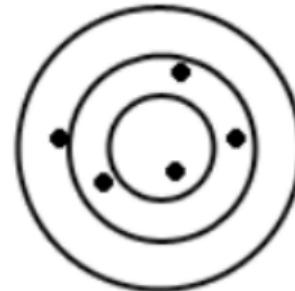
Data

Anecdote: it doesn't accumulate to be data.

Data



High bias, low variance



Low bias, high variance



High bias, high variance



Low bias, low variance

Data

Features

Feature Engineering

Data

One of K = one-hot encoding

Data

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

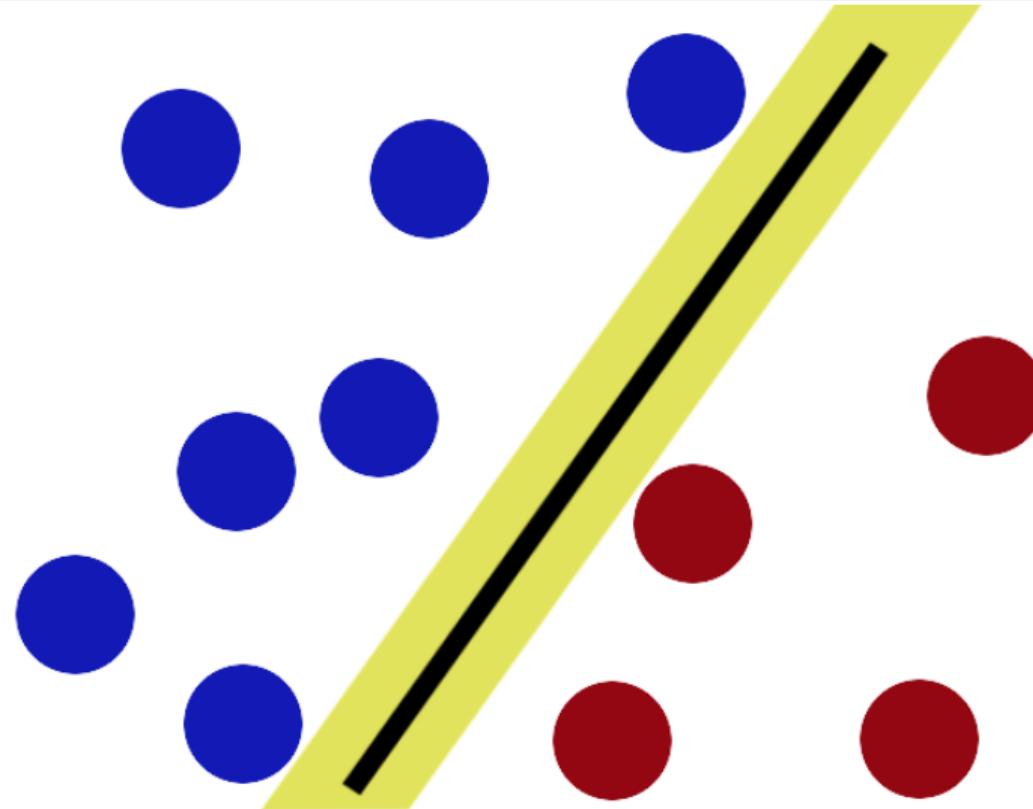
- Set aside (partition) data for testing (e.g., 70% / 30%)
- Learn on training set, test on testing set
- When searching hyperparameters, set aside again (e.g., 60% / 20% / 20%)

SVM

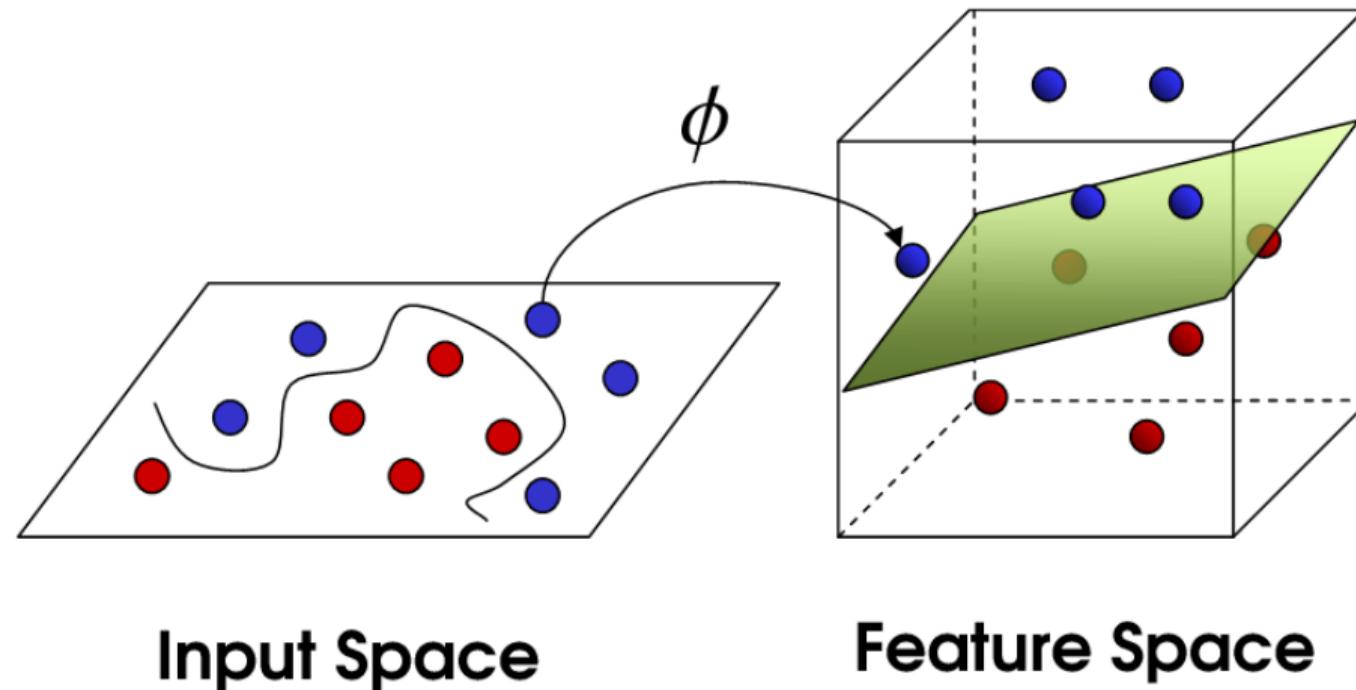
?

copperking@reddit

SVM



SVM



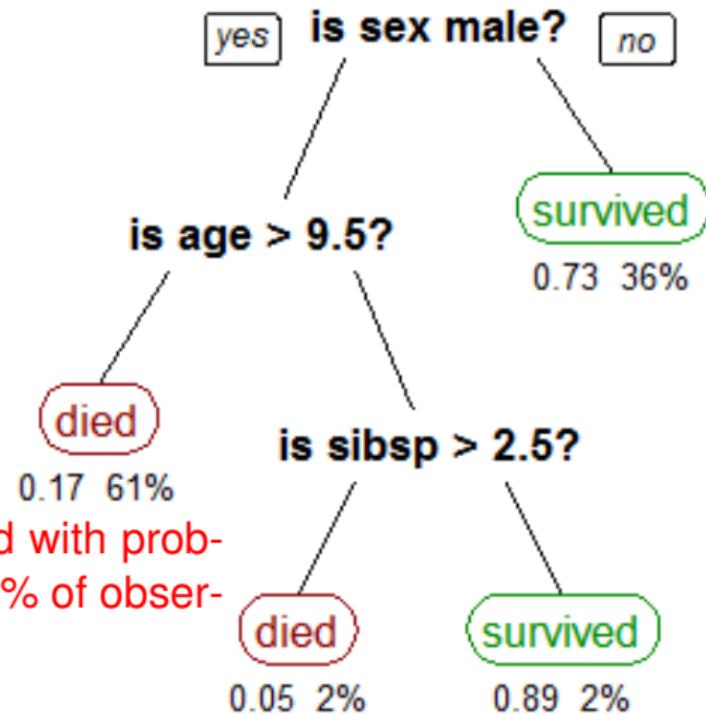
CART

?

Random Forest

?

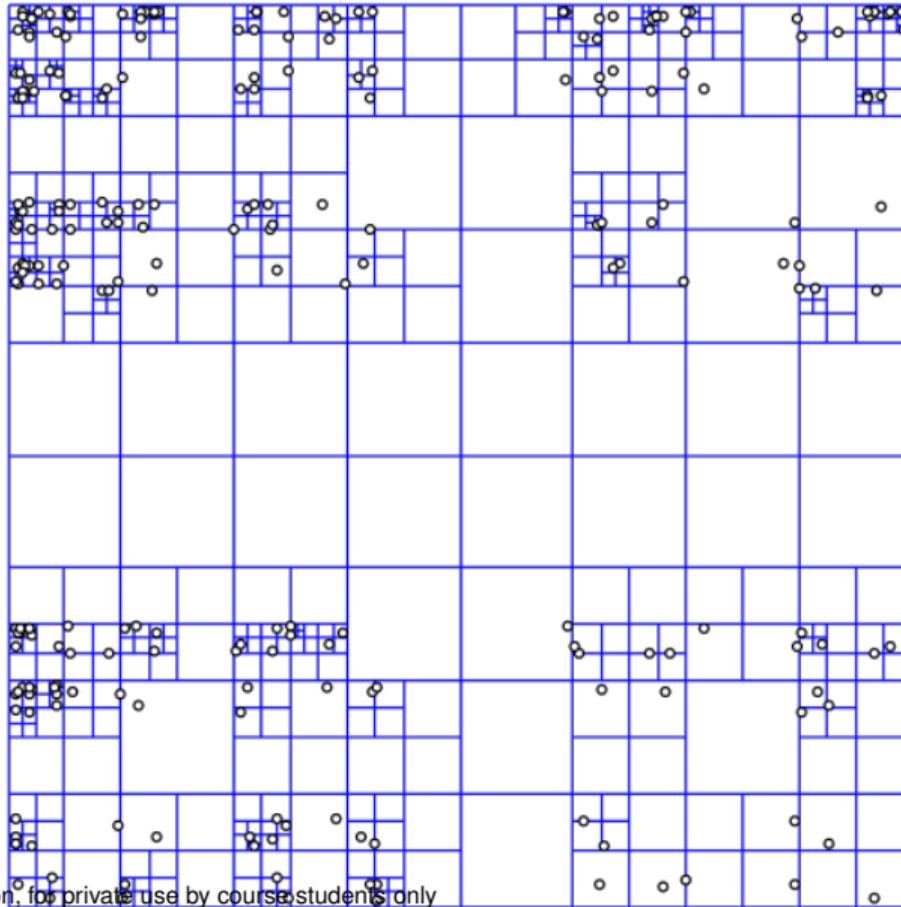
Decision Trees



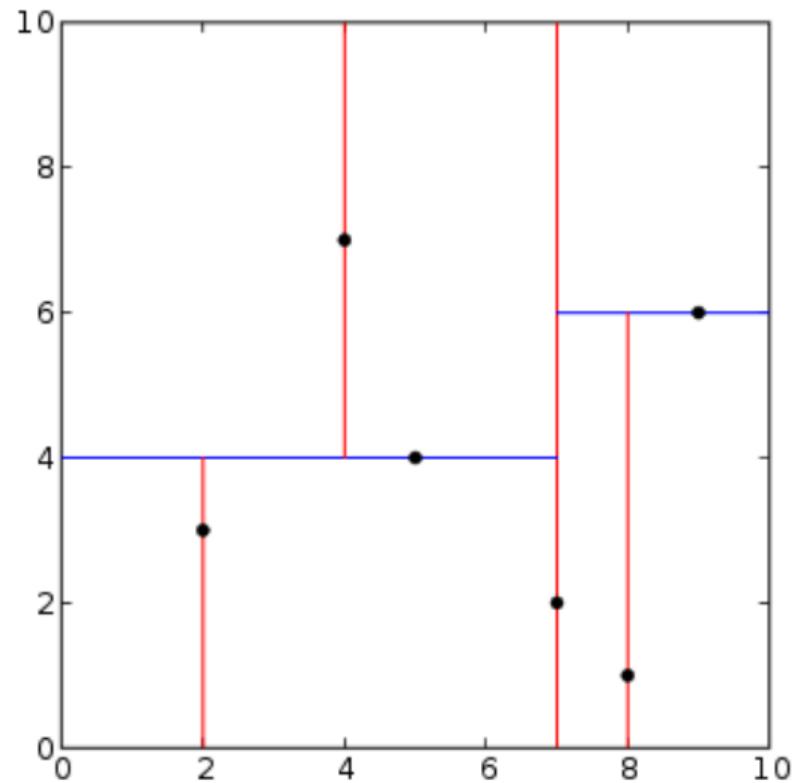
E.g., passengers died with probability .17 which is 61% of observations

Stephen Milborrow

Quadtree



kd-tree



Random subspace methods

Draw k hyperplanes at random.

- With text, dot product, so random planes through origin
- With nearly everything else, random planes period

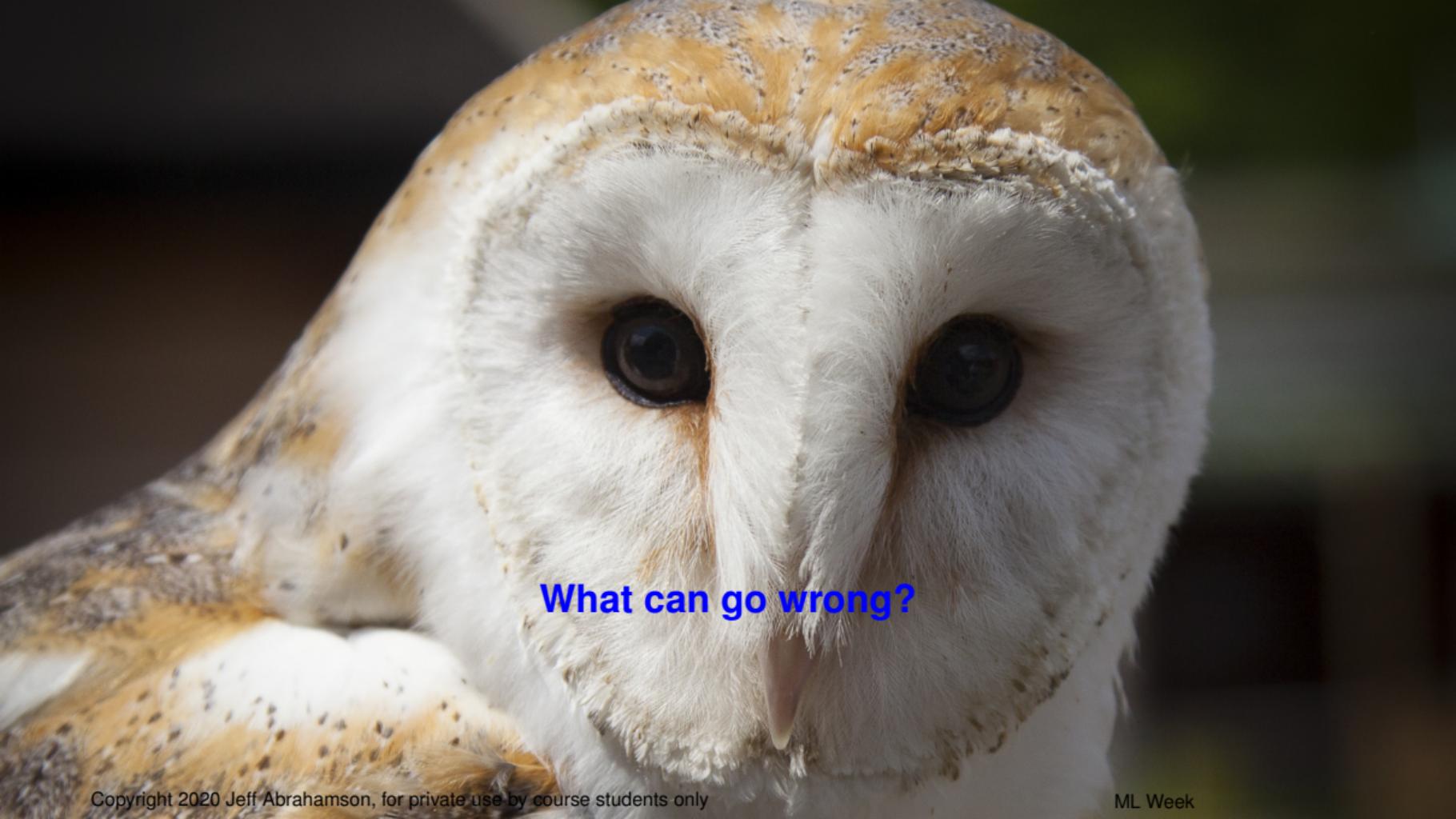
Result: $\log(n)$ search time

Decision Trees

Variations

- Classification tree
- Regression tree

CART = classification and regression trees

A close-up photograph of a Barn Owl's face. The owl has large, dark, almond-shaped eyes and a white, feathered facial disc with a distinct dark border. Its plumage is a mix of light brown and white with dark spots. The background is dark and out of focus.

What can go wrong?

Decision Trees

Ensemble methods

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

Bootstrap aggregating = bagging

Bagging

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

A type of model averaging (ensemble method).

Bootstrap aggregating = bagging

- Training set D of size n
- Sample D *with replacement* to create D_1, \dots, 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](#)

A photograph of a dense forest in a thick fog. The trees are tall and thin, their trunks silhouetted against the light. Bare branches reach out from the left side of the frame. In the center, the word "questions?" is written in a bold, blue, sans-serif font.

questions?

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

Copyright 2020 Jeff Abrahamson, for private use by course students only

ML Week



Eddie Bell @ Lyst

Copyright 2020 Jeff Abrahamson, for private use by course students only

ML Week



Eddie Bell @ Lyst

Copyright 2020 Jeff Abrahamson, for private use by course students only

ML Week



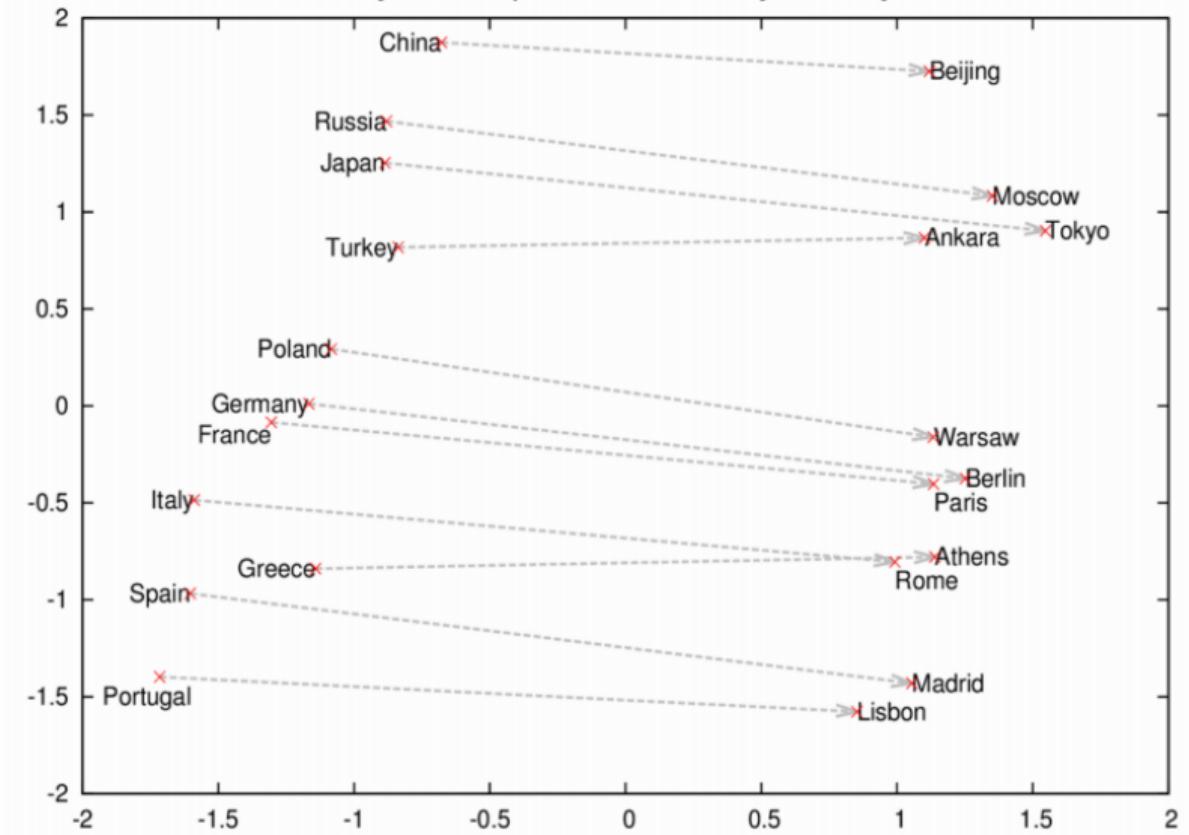
-1	0	+1
-2	0	+2
-1	0	+1



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



Country and Capital Vectors Projected by PCA



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

Similar to 'dress'



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?

Copyright 2020 Jeff Abrahamson, for private use by course students only

ML Week

A black and white photograph of a dark, graffiti-covered hallway. A single chair sits in the center of the floor. The ceiling is made of dark, textured panels with some light fixtures. The walls are covered in graffiti, including the number "25" on one wall. The overall atmosphere is gritty and industrial.

questions?

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...

In addition to PCA...

There are newer methods that are sometimes better.

- Sometimes they are faster.
- Sometimes the projection makes more sense.

t-SNE is the most common you'll encounter.



questions?