

Machine Learning

SVM and Random Forests

Jeff Abrahamson

octobre 2020

Review

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

- 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

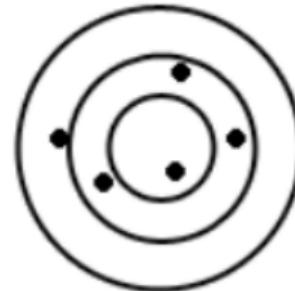
Data

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

Easy Features

We'll see more

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.

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

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

- 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

What I described yesterday:

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

The missing point: 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. 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

- 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

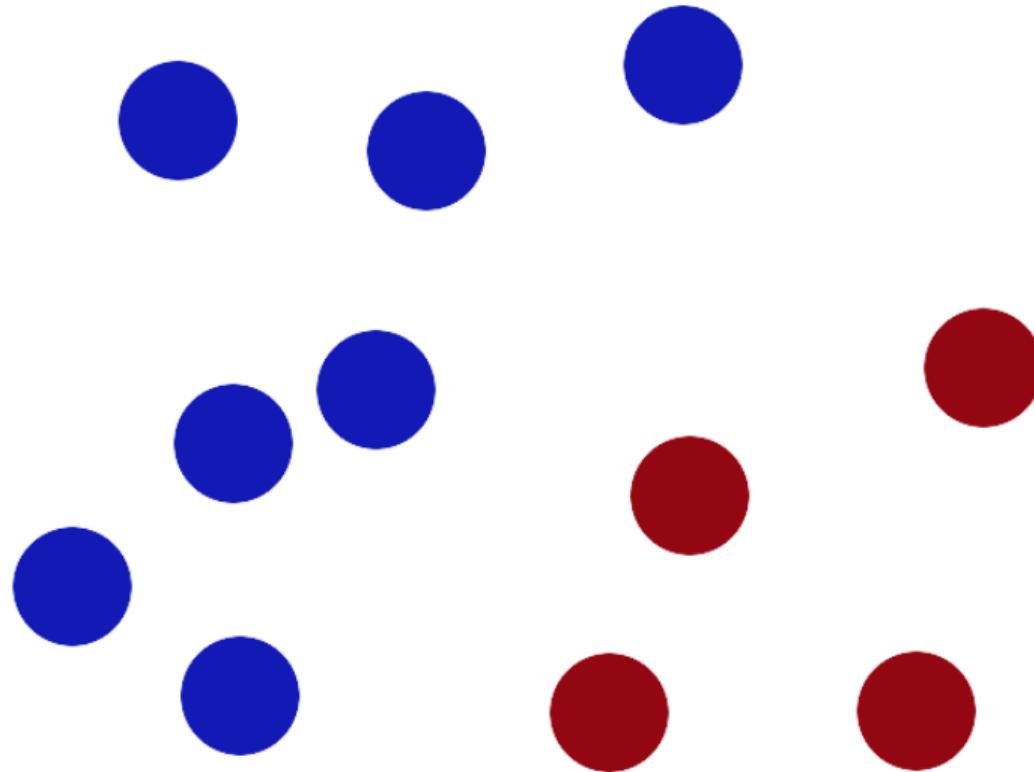
Testing

- 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%)

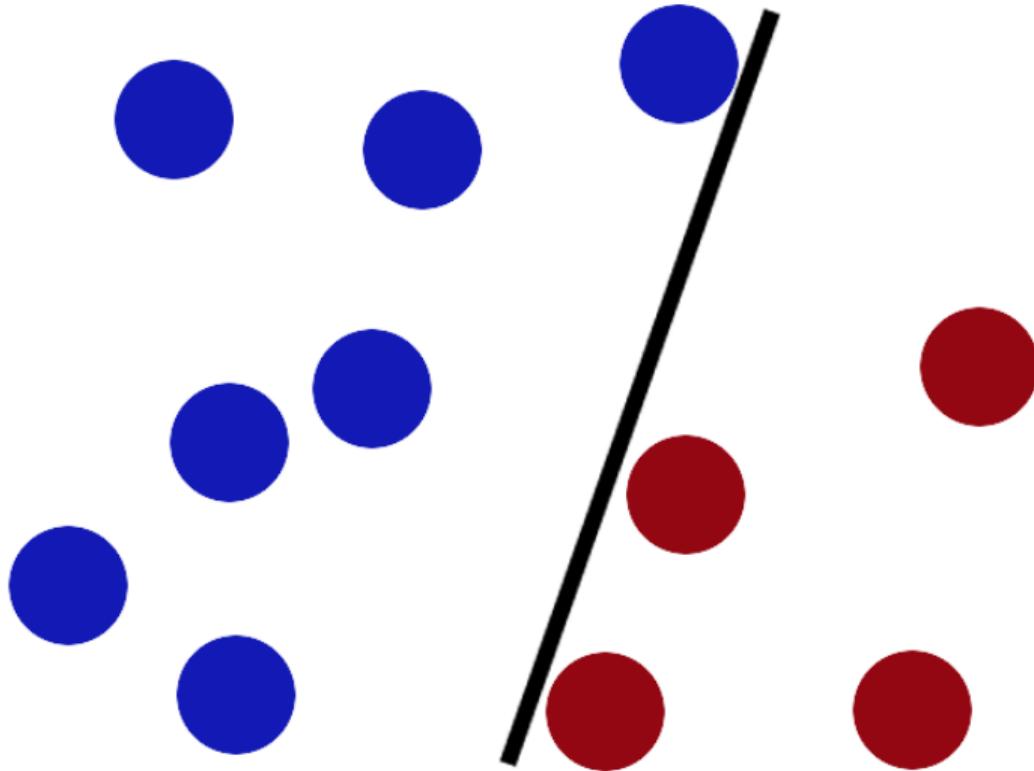
A wide-angle photograph of a mountainous landscape. In the foreground, a calm lake reflects the surrounding environment. On either side of the lake are steep mountains covered in dense green forests. The sky above is filled with scattered white and grey clouds.

questions?

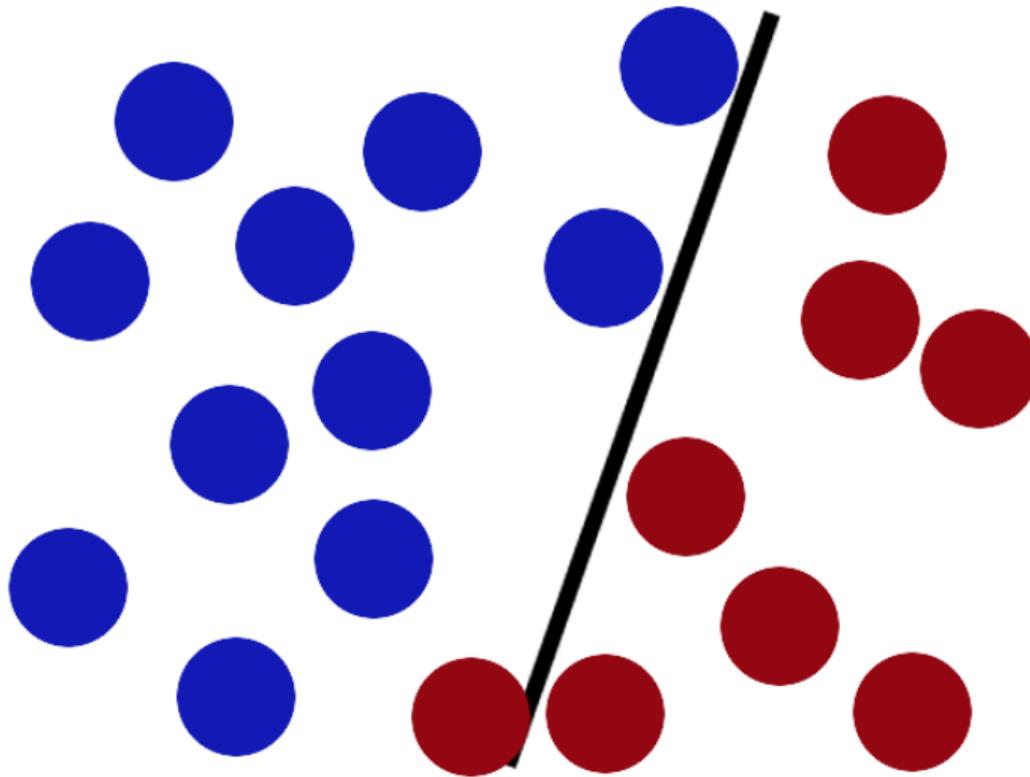
The simple explanation



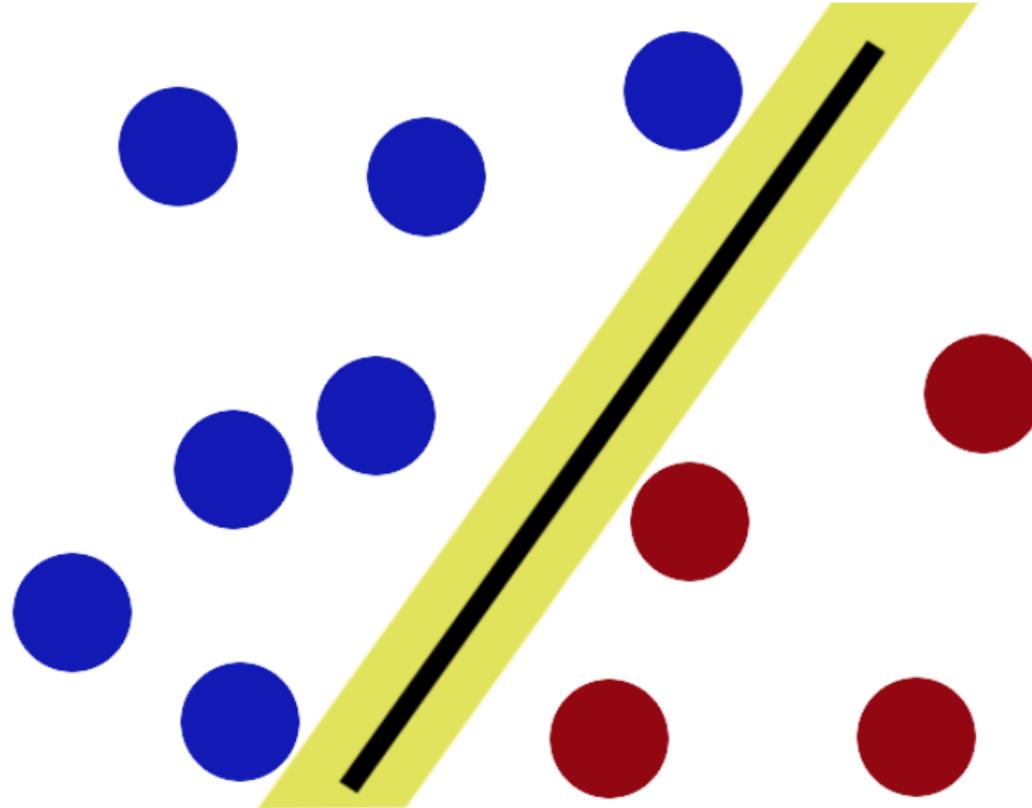
The simple explanation



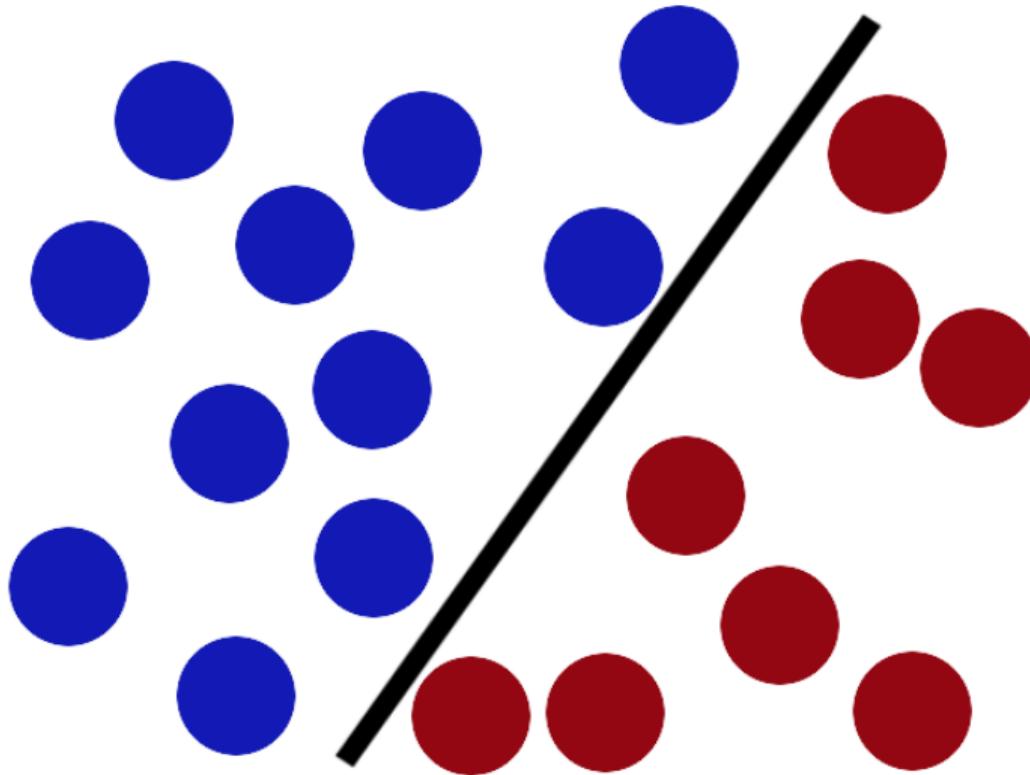
The simple explanation



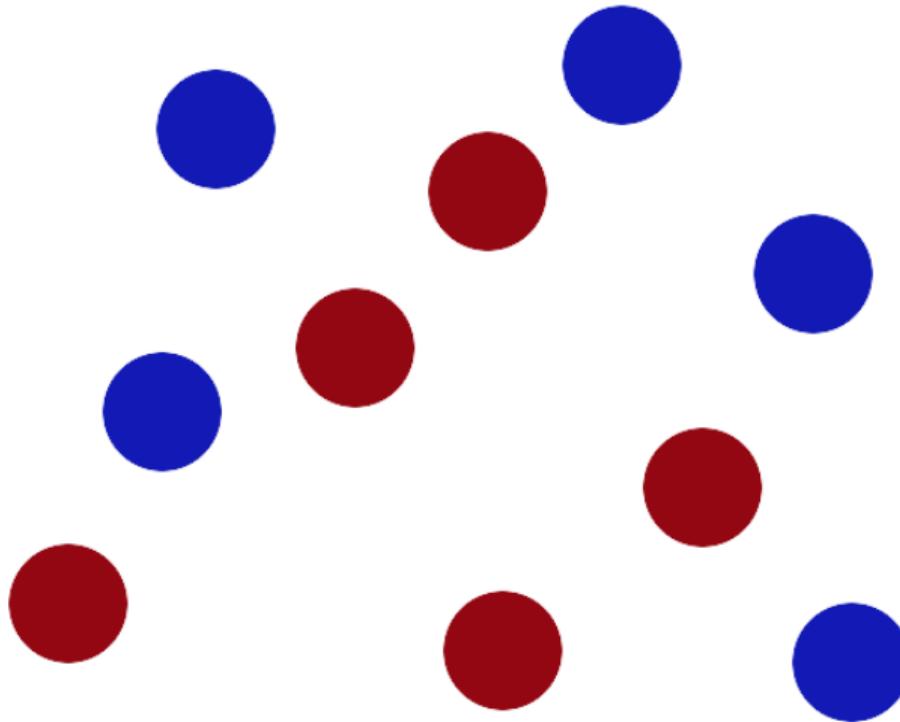
The simple explanation



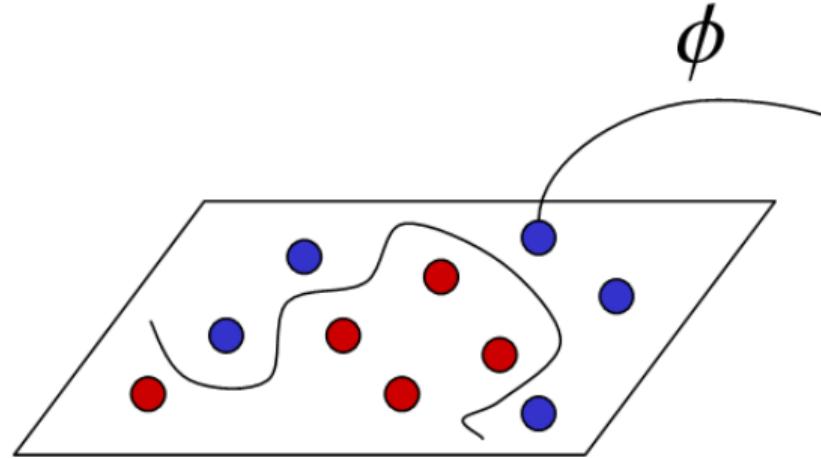
The simple explanation



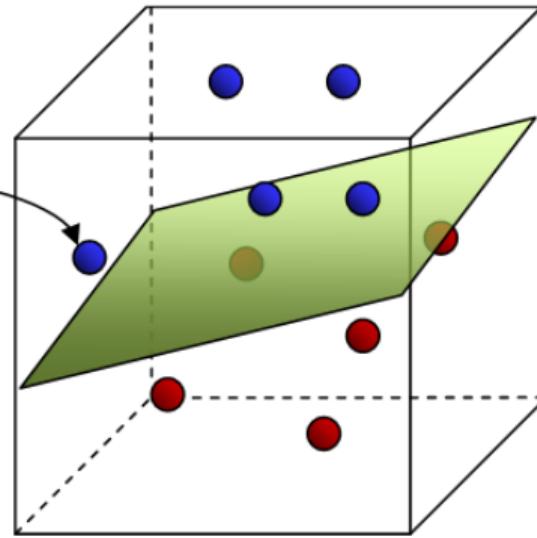
The simple explanation



The simple explanation

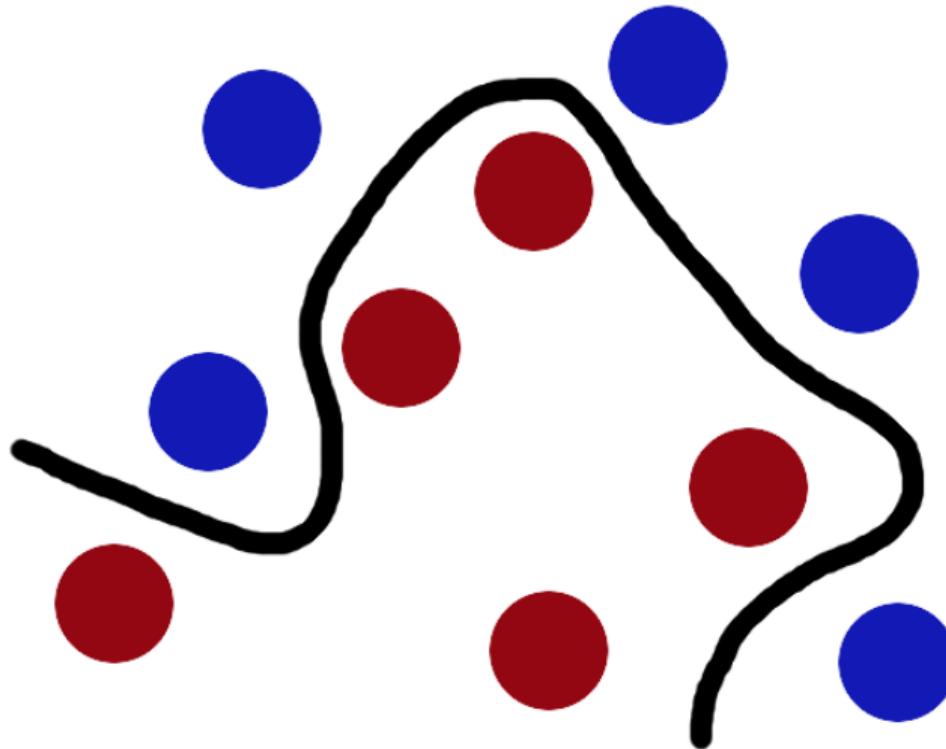


Input Space



Feature Space

The simple explanation



video time

questions?

other's potential
not proceed

and can affect

the patient's mental

other's potential

not proceed

in practice

of a talk

with a patient

and suggest

to take a

longer time

and suggest

to take a

longer time

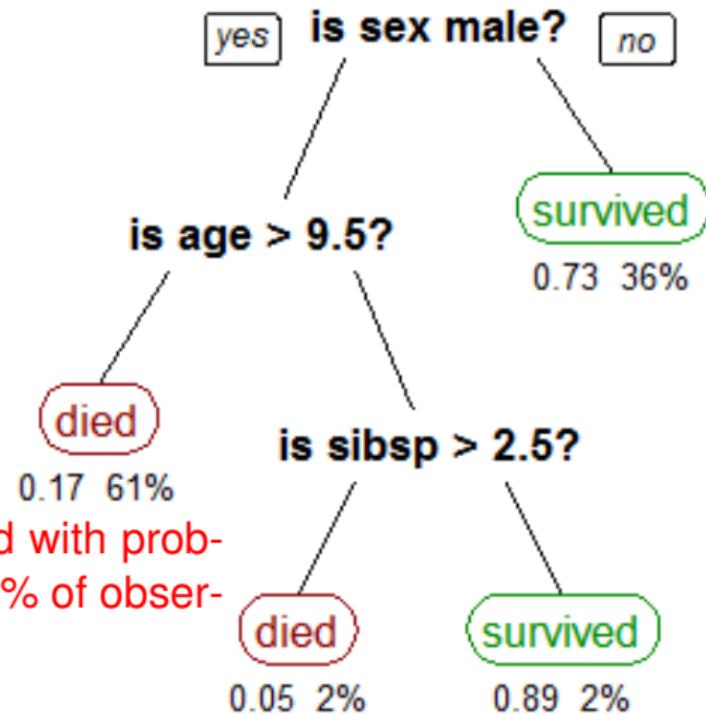
and suggest

to take a

longer time



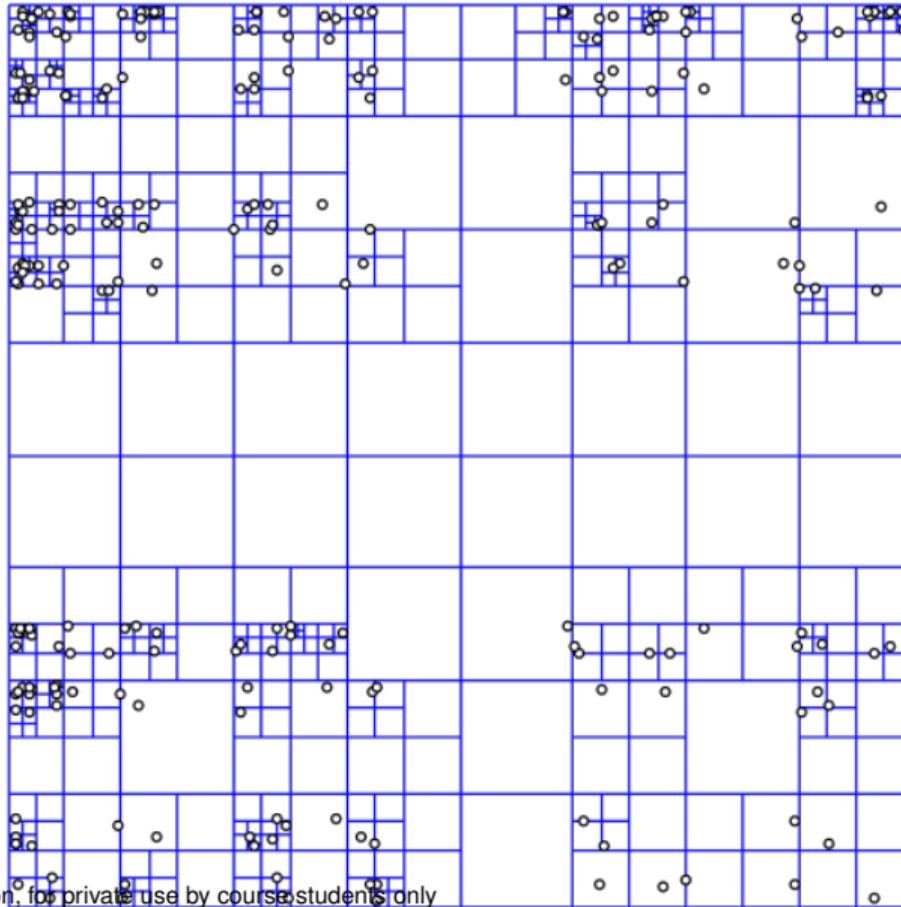
Decision Trees



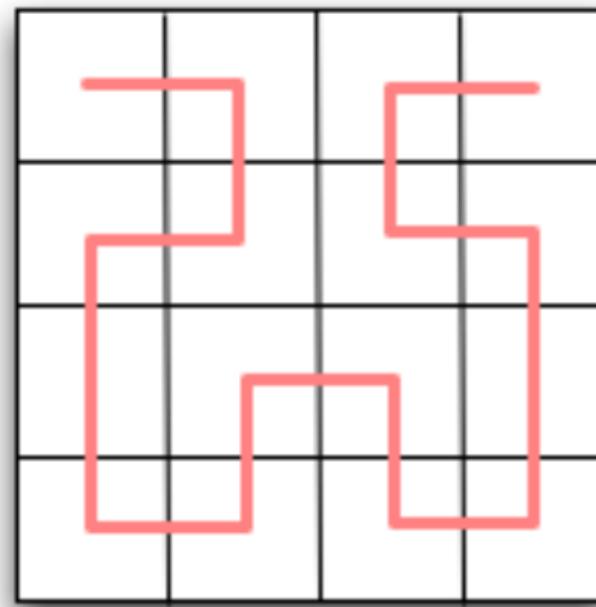
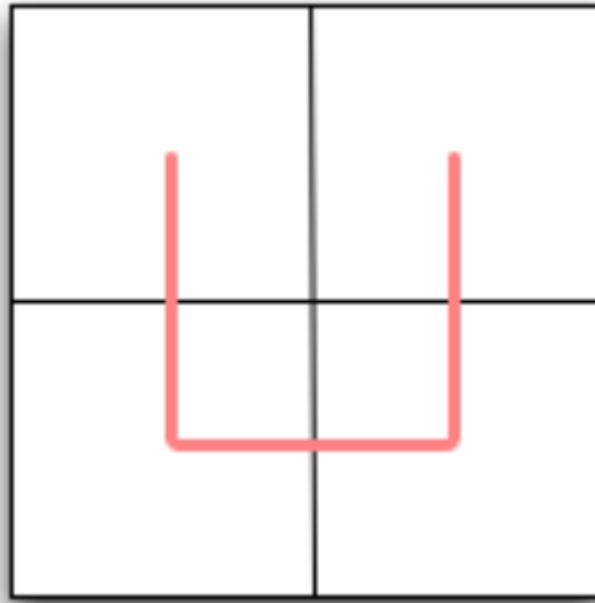
E.g., passengers died with probability .17 which is 61% of observations

Stephen Milborrow

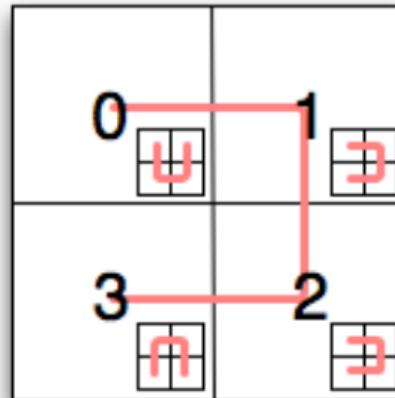
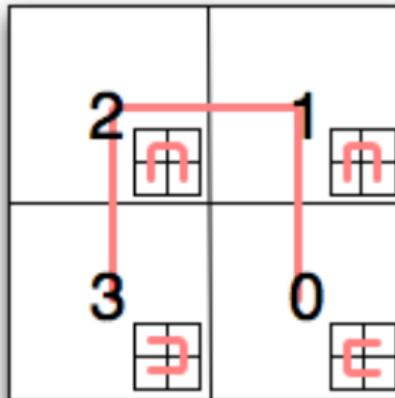
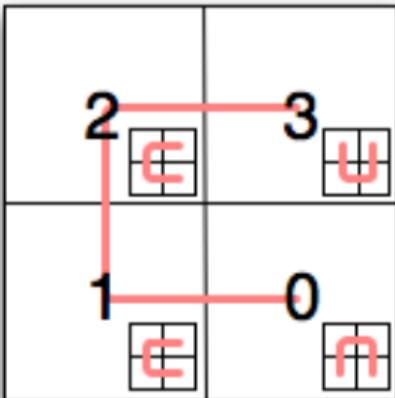
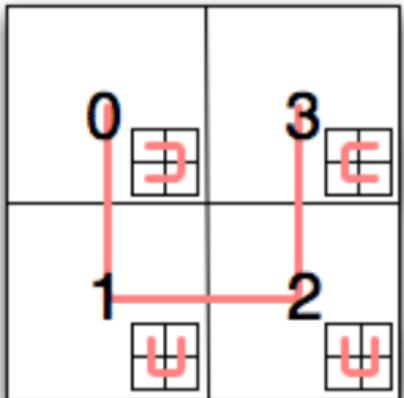
Quadtree



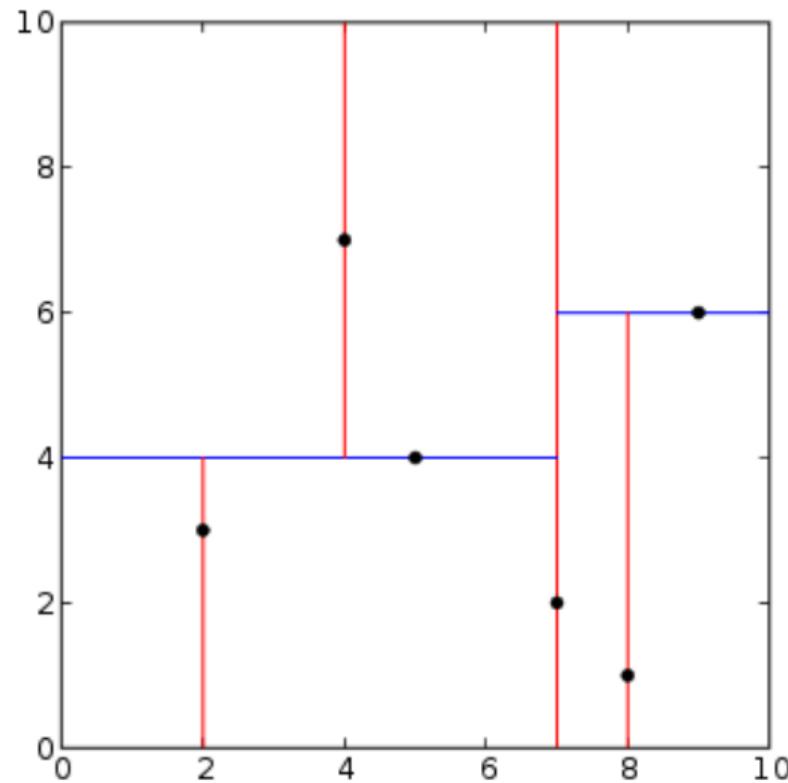
Quadtree



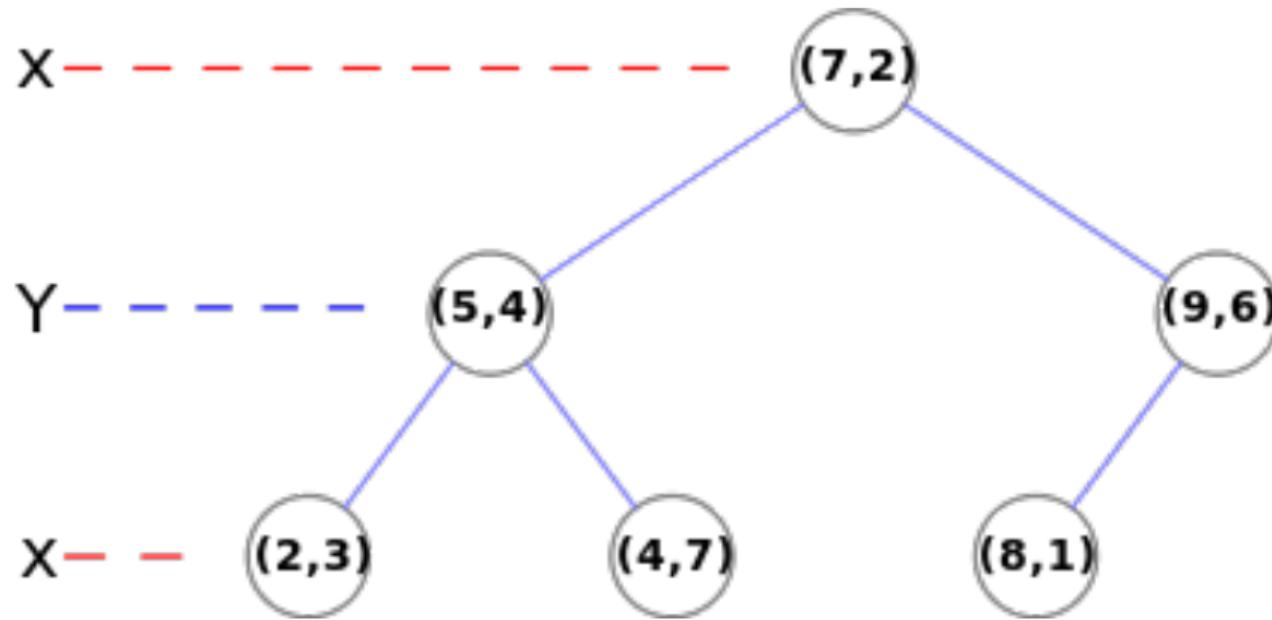
Quadtree



kd-tree



kd-tree



kd-tree

[***kd*-tree animation**](#)

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 'M' shape. Its plumage is a mix of light brown and white with dark spots. The background is blurred.

What can go wrong?

Decision Trees

Ensemble methods

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

Bootstrap aggregating = bagging

Bootstrap

A family of statistical methods using sampling with replacement.

(Also: an example of an ensemble method.)

Bootstrap aggregating = 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

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)

Bootstrap aggregating = bagging

Do not confuse with

- Boosting (and AdaBoost)
- Bootstrap (statistics)
- Cross validation

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. The scene is filled with tall, thin trees standing in a grid-like pattern. Their trunks are mostly bare, with some having small, dark, curved branches. The ground is covered in a mix of green moss and fallen brown leaves. In the background, the forest continues through a thick layer of fog, which obscures the sky and creates a mysterious atmosphere.

questions?