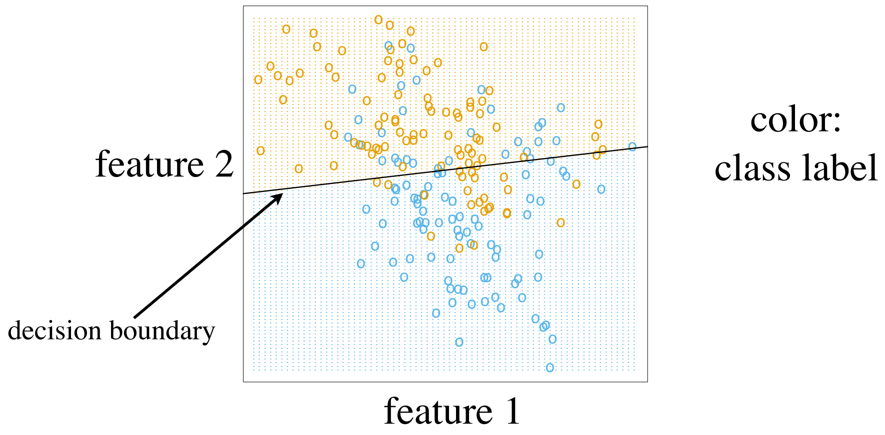
Decision Trees

Charles Sutton
Data Mining and Exploration
Spring 2012

The Classification Problem



feature 2

feature 1

decision boundary

color:
class label

Figure from [Hastie, Tibshirani, and Friedman, 2009]

The Classification Problem

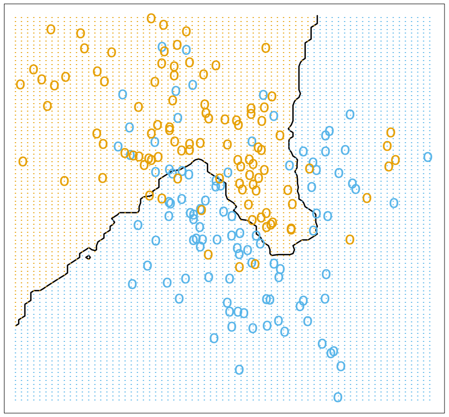


Figure from [Hastie, Tibshirani, and Friedman, 2009]

Classification Methods

- Naive Bayes
- Logistic Regression
- Decision Trees
- Nearest Neighbour
- Neural Networks
- Support Vector Machines
- Ensemble Methods

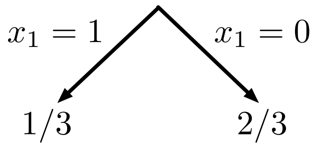
Classification Methods

- Naive Bayes
- Logistic Regression
- **Decision Trees** (CART)
- Nearest Neighbour
- Neural Networks
- Support Vector Machines
- **Ensemble Methods** (Bagging, Boosting)

Decision Trees

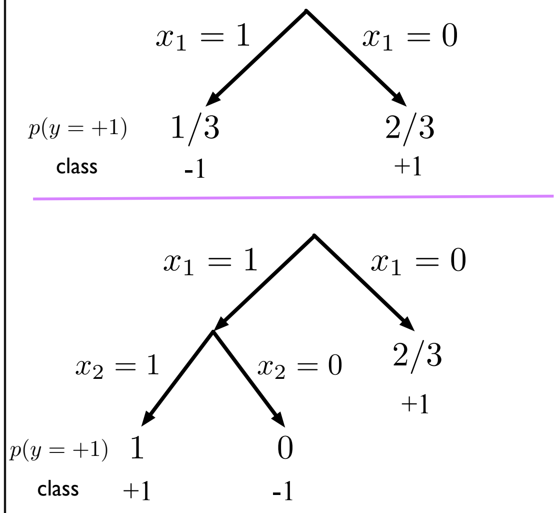
- This will be very fast
- For a refresher see IAML lecture video
  - http://groups.inf.ed.ac.uk/vision/VIDEO/2011/iaml.htm (lecture 5)
- (or look at readings)

What Decision Trees Look Like



y	x_1	x_2	x_3
1	1	1	0
-1	1	0	1
-1	0	0	1
1	0	0	1
1	0	1	1
-1	1	0	0

What Decision Trees Look Like



y	x_1	x_2	x_3
1	1	1	0
-1	1	0	1
-1	0	0	1
1	0	0	1
1	0	1	1
-1	1	0	0

How to build trees

- First idea: Find a tree that is always correct on training data
- Problem: This idea is stupid.

How to build trees

- Second idea: Find the smallest possible tree that fits the training data
- This doesn't work either.

How to build trees

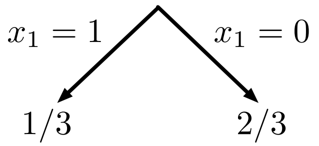
Solution:

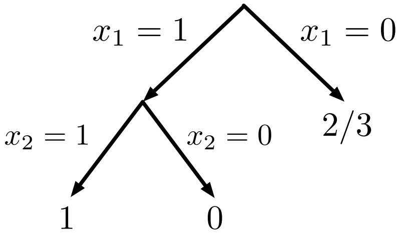
- Be recursive.
- Be greedy.

Tree Building Algorithm

```
Start with tree containing only root
Assign all instances to the root
Repeat:
    Pick a leaf $v$ in the tree
    If no features left, ignore $v$
    If all instances have same class, ignore $v$
    Choose a feature $x_j$ to split the tree on
    Add children to $v$, one for each value of $x_j$
      Subdivide instances of $v$ accordingly
Until all leaves have been processed
```

What Decision Trees Look Like

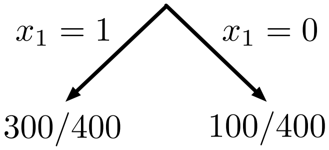




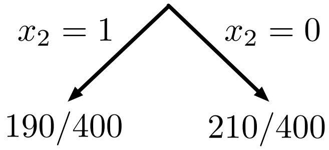
y	x_1	x_2	x_3
1	1	1	0
-1	1	0	1
-1	0	0	1
1	0	0	1
1	0	1	1
-1	1	0	0

How to choose features to split?

- Basically need a measure of the "purity" of instances at a leaf



GOOD

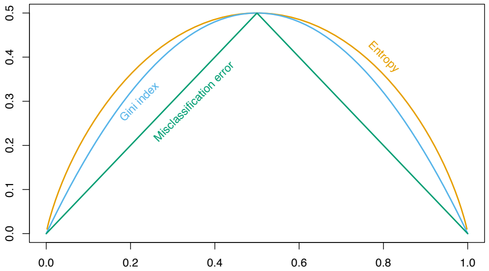


BAD

How to choose features to split?

Gini $p_{m,-1} p_{m,1}$

Cross-entropy $p_{m,-1} \log p_{m,-1} + p_{m,1} \log p_{m,1}$



Extensions

- Multiple classes
- Continuous values
- Pruning

Advantages, disadvantages

- Good: Fast to train, Easy to interpret
- Bad: Accuracy not great, Unstable

Readings

Examinable readings:

- Section 9.2 of Hastie, Tibshirani, and Friedman
  - http://www-stat.stanford.edu/~tibs/ElemStatLearn/download.html
- HMS Section 10.5
- Also see IAML Lecture video earlier