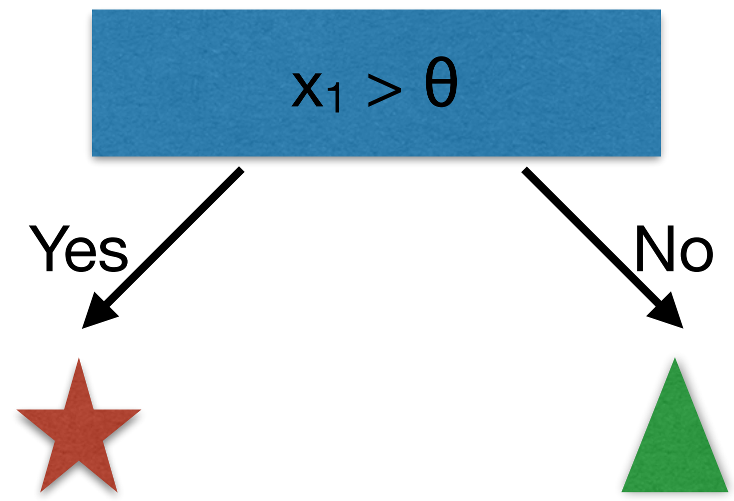
Machine Learning for Applications in Computer Vision

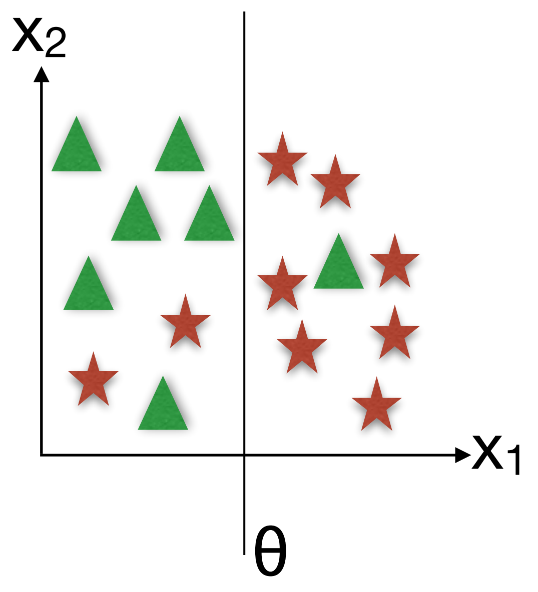
Tree-based Classifiers

Decision Stump

- One level decision tree
- One internal node (root) connected to its terminal nodes (leaves)

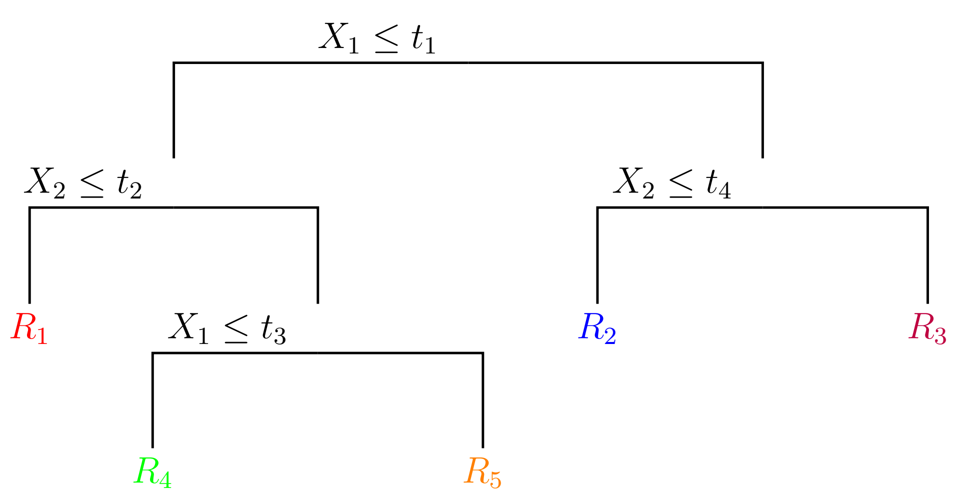


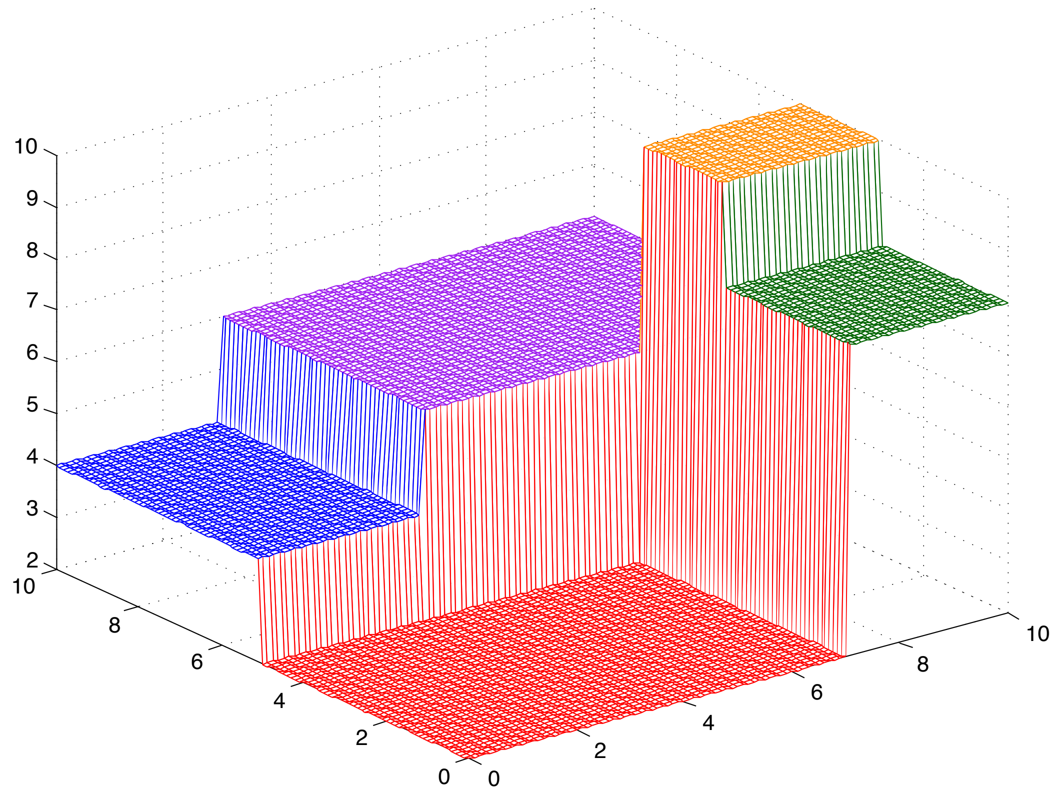
- **Goal:**
  Find axis aligned hyper plane that minimises the class. error
- Class. error is always better than random guessing (0.5)



Decision Trees

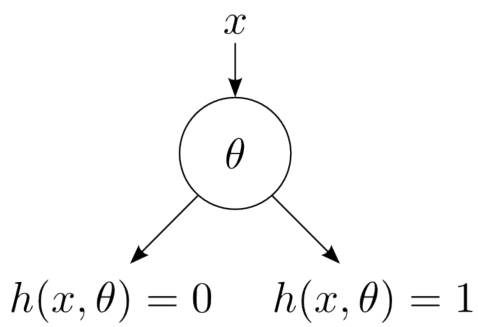
- Classification and Regression Trees (CART)
- Extension of Decision Stump
- Partition the input space recursively
- Define a label for each resulting region of the input space.



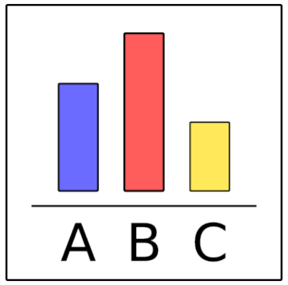


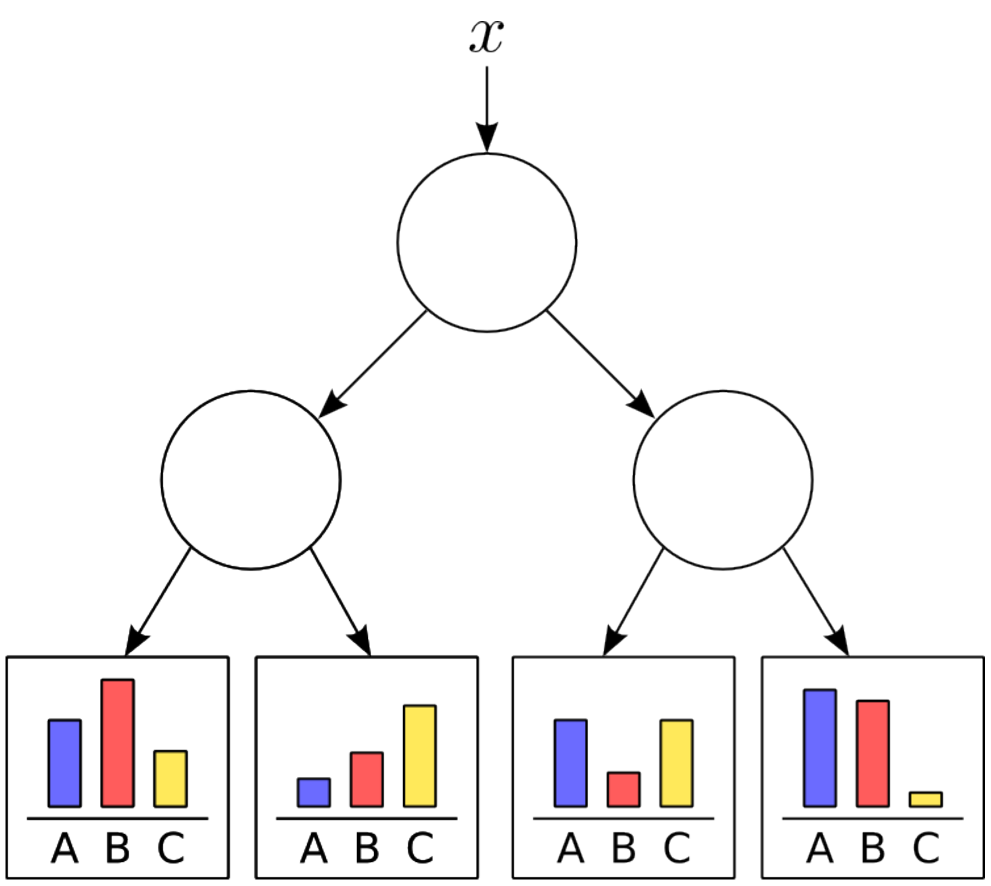
Decision Trees

- Regression: assign mean response to each leaf (piecewise constant surface)
- Classification: store the distribution over class labels in each leaf
  - Inner node:



  - Leaf node:





Growing a Tree

- NP-Complete problem (NP: Non-deterministic Polynomial time)
- Solution is locally optimal

- Minimise a cost function to find the best feature and its best threshold on each node
- Split the data on each node based on the chosen feature and the threshold
- Stopping criteria for growing the tree
  - reduction of cost too small ?
  - maximum depth is reached ?
  - is the distribution in the subtrees homogenous ? (pure dist.)
  - is the number of samples in the subtrees too small ?

Growing a Tree

- Regression cost:

$$\mathrm{cost}(D) = \sum_{i \in D} (y_i - \bar{y})^2 \qquad \bar{y} = \frac{1}{|D|} \sum_{i \in D} y_i$$

- Classification cost:
  - Misclassification rate:

$$\frac{1}{|D|} \sum_{i \in D} |y_i \neq \hat{y}|$$

  - Entropy:
    - same as maximising the information gain

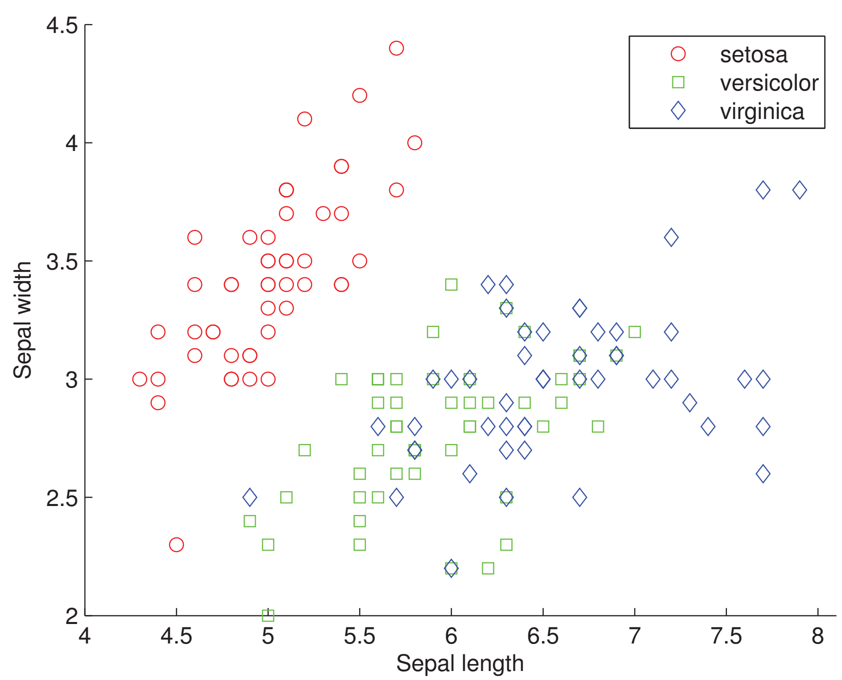
$$H(\hat{P}) = -\sum_{c=1}^{C} \hat{P}_c \log \hat{P}_c$$

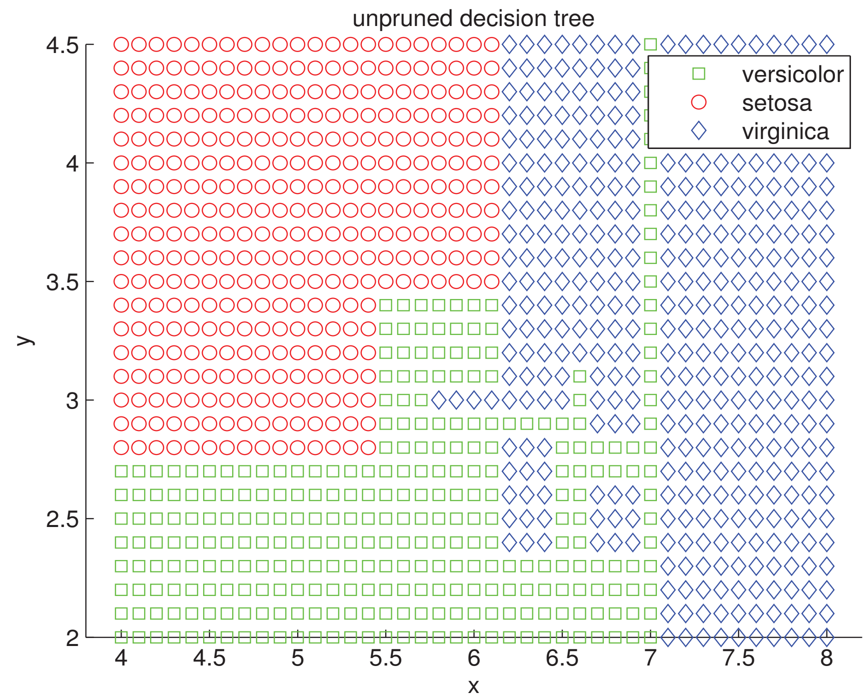
  - Gini Index:
    - expected error rate

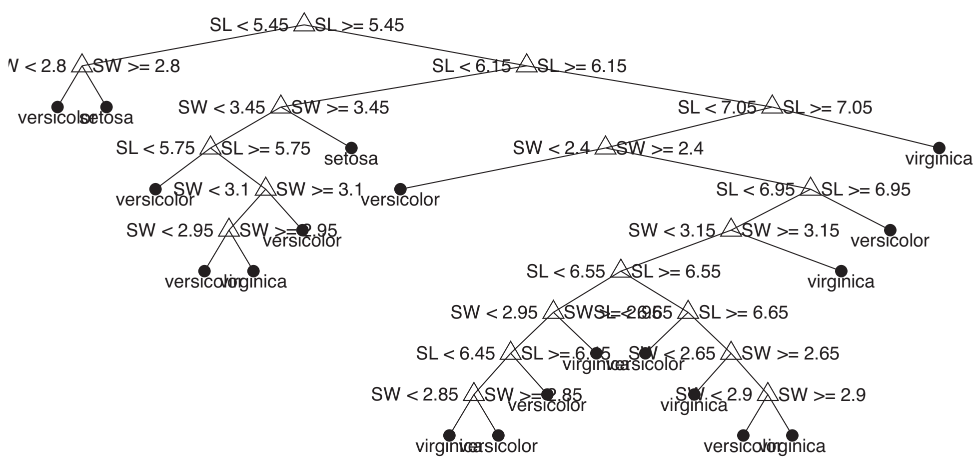
$$1 - \sum_{c} \hat{P}_c^2$$

Pruning a tree

- Growing a tree too large yields overfitting
- Solution: build a full tree and then prune it

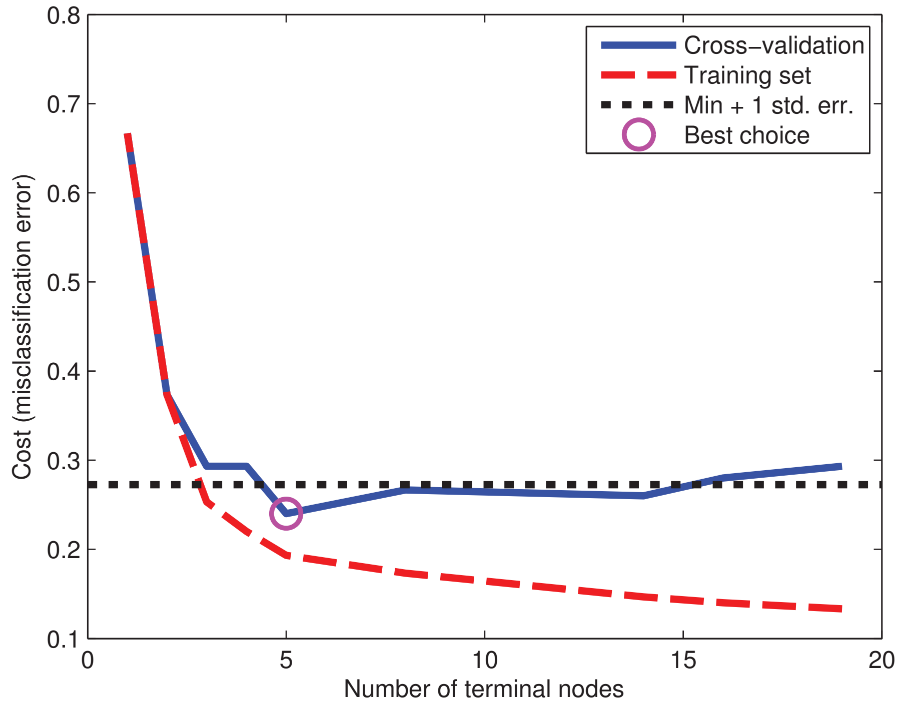


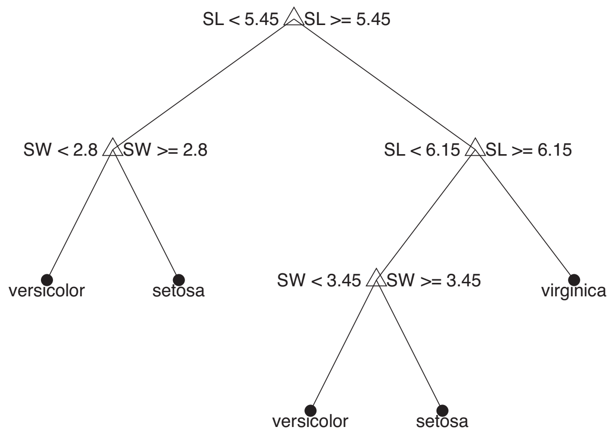


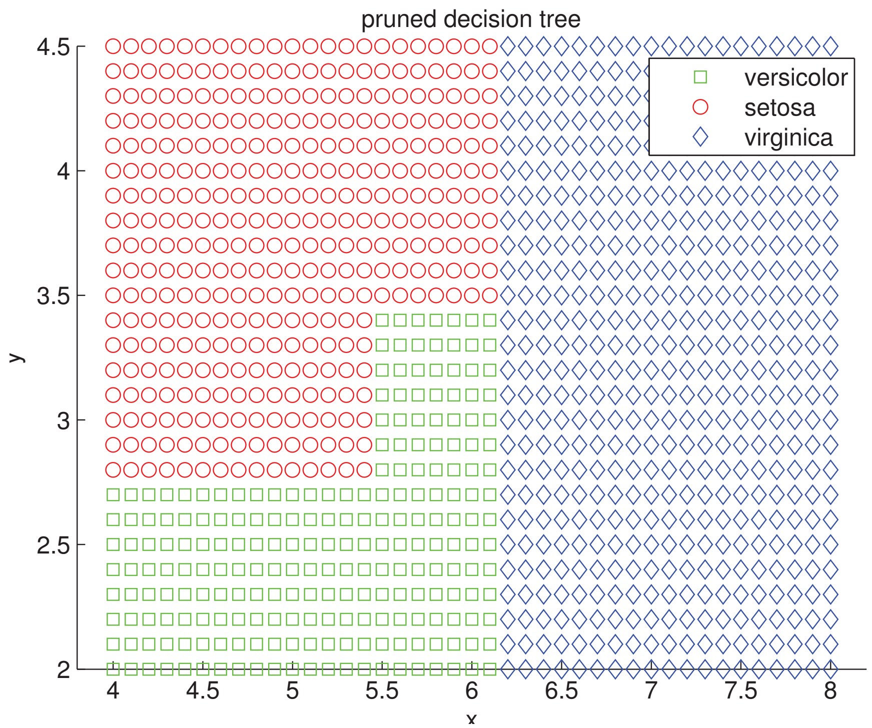


Pruning a tree

- Growing a tree too large yields overfitting
- Solution: build a full tree and then prune it







Pros (CART)

- easy to interpret
- can handle mixed discrete and cont. data
- insensitive to monotone transformations
- CART perform automatic variable selection
- relatively robust to outliers
- scale well to large datasets
- can be modified to handle missing inputs

Cons (CART)

- DO NOT predict very accurately
  - due to the greedy training procedure
- Trees are unstable
  - small change in the input might yield a large effect on the tree structure
- Trees are high variance estimators
  - Solution: Random Forests

Random Forests

- Reduce the variance of estimate by
  - Train M trees on different subsets of the data:

$$f(x) = \sum_{m}^{M} \frac{1}{M} f_m(x)$$

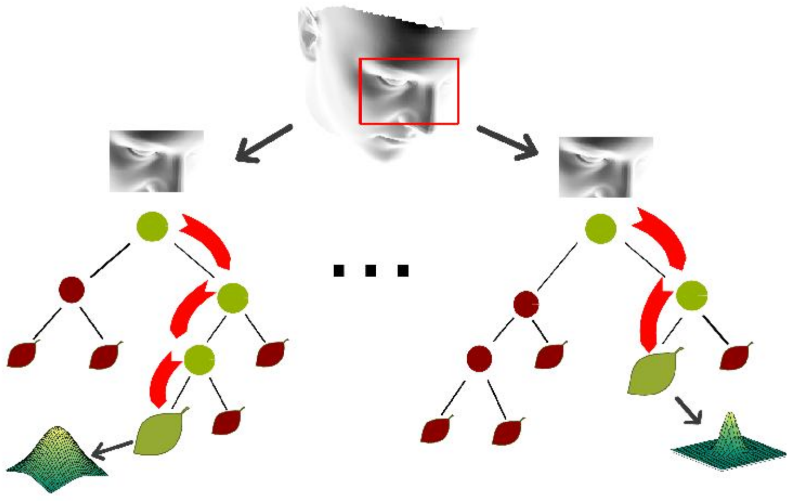
    - !!! highly correlated predictors
  - Solution: Choose data as well as variable (feature) randomly
- Known as Random Forests. RF has a high accuracy and widely used in practical studies.

Random Forests

- **Real-Time Object Segmentation with Semantic Texton Forests**
  - James Shotton (winner of CVPR 2008 Demo Prize)

Random Forests

- **Real Time Head Pose Estimation with Random Regression Forests**
  - Fanelli et al. (CVPR 2011)



- https://www.youtube.com/watch?t=136&v=sxUkGGGtRBU