VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wananga o te Upoko o te Ika a Maui

School of Engineering and Computer Science

COMP 307 — Lecture 06

Machine Learning 3

Decision Tree Learning Method

Dr Bing Xue

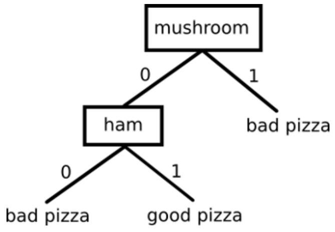
bing.xue@ecs.vuw.ac.nz

Outline

- Decision tree learning vs learned decision trees
- How to build a decision tree using a set of instances
- How to measure a DT node: (im)purity measures
- Numeric attributes: splitting points

Decision Tree

Jane's Pizza Party



http://www.aispace.org/exercises/exercise7-a-1.shtml

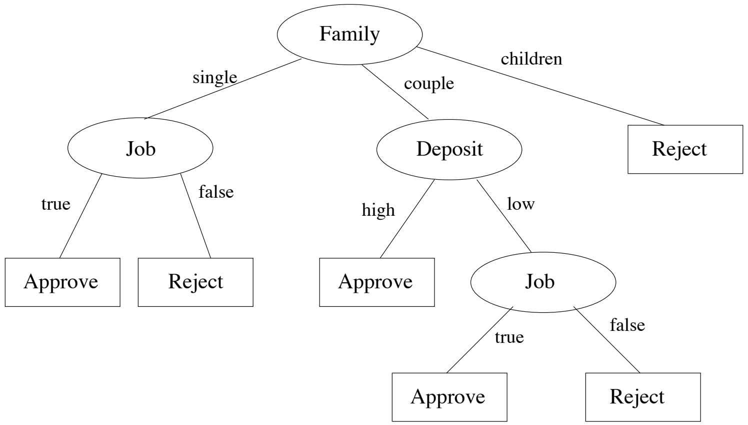
Decision Trees vs Decision Tree Learning

- Representation of the classifier: *decision trees*
  - An alternative representation for classifiers – Symbolic, not probabilistic
  - “Easier” to interpret
- Learning process: *decision tree learning (method)*
  - Specify a procedure for deciding a class
  - One of the oldest classification learning methods in AI
  - Also developed independently in Statistics/Operations Research
- Decision tree = decision tree learning?

Example (Training) Data Set

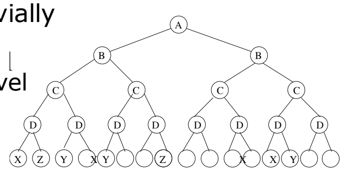
	Job	Deposit	Family	Class
A	true	low	single	Approve
B	true	low	couple	Approve
C	true	low	single	Approve
D	true	high	single	Approve
E	false	high	couple	Approve
1	true	low	couple	Reject
2	false	low	couple	Reject
3	true	low	children	Reject
4	false	low	single	Reject
5	false	high	children	Reject

Decision Trees



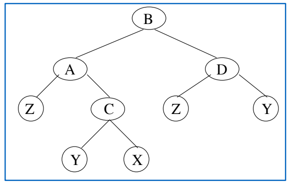
Building Decision Trees

- You can always build a decision tree trivially
  - Choose some order on the attributes
  - Build tree with one attribute for each level
  - Label each leaf with appropriate class



- Problems
  - Each leaf represents a possible instance
  - All we are doing is remembering every instance — no generalisation, no prediction, no learning

- Solution
  - Find a small decision tree
  - capture the common characteristics of instances
  - probably generalise to predict classes for unseen instances



Building A Good Decision Tree

- Input: Instances
- Output: a “good” decision tree classifier
  - Critical issue: choosing which attribute to use next

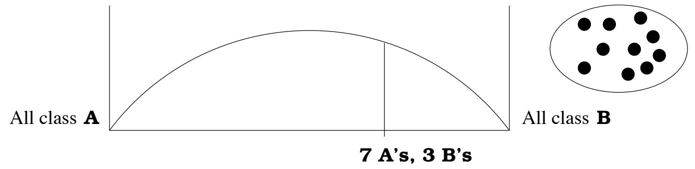
- DT algorithm

```
Examine set of instances in the root node

If set is "pure" enough, or no more attributes
      then stop
Else
      Construct subsets of instances in the subnodes
      Compute average "purity" of subnodes
      Choose the best attribute

Recurse on each subnode
```

Measuring Purity

- Need a measure of how "pure" a node is
  - all one class −→ pure −→ can predict the class
  - mixture of classes −→ impure −→ have to ask more questions
- Several functions
  - probability based
  - information theory based
- Choose the attribute whose children have the best purity

(Im)Purity Measure: P(A)P(B)



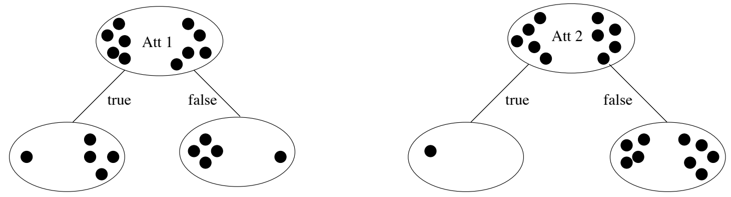
- Impurity:
  - $P(A)P(B) = \frac{m}{m+n} + \frac{n}{m+n} = \frac{m*n}{(m+n)^2}$
  - m: number of A's; n: number of B's
- Goodness of attribute:
  - average impurity of subnodes

smaller the better

Weighting the Impurities

- How do we take the average?

smaller the better



impurity = 1/5 * 4/5 = 16%

impurity = 4/5 * 1/5 = 16%

average = 16%

impurity = 1/1 * 0/1 = 0%

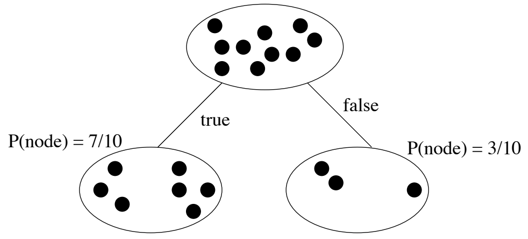
impurity = 4/9 * 5/9 = 24.6%

average = 12.3% ?

average = 20% ?

- Need to weight the nodes by probability of going to node:

Weighting the Impurities (Continued)



impurity(node) = 4/7 * 3/7 = 12/49

impurity(node) = 2/3 * 1/3 = 2/9

- Goodness of attribute = weighted average impurity of subnodes

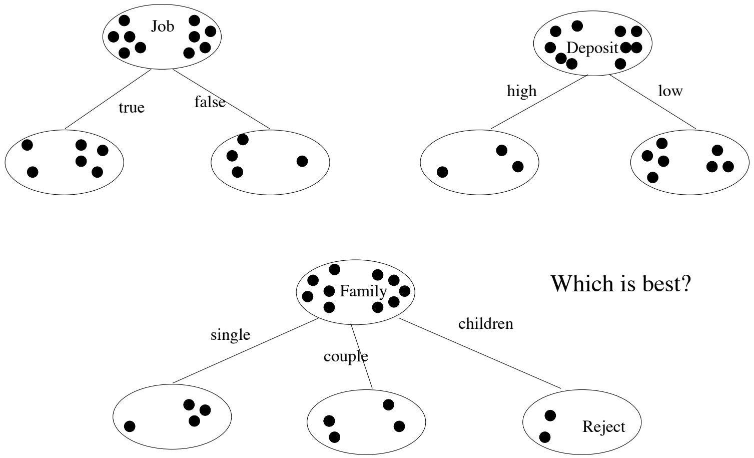
$=\sum_i [P(node_i) * impurity(node_i)]$

= (7/10 × 12/49) + 3/10 × 2/9

= 84/490 + 6/90 = 0.238

Decision Tree Building

- Choose the best attribute



Which is best?

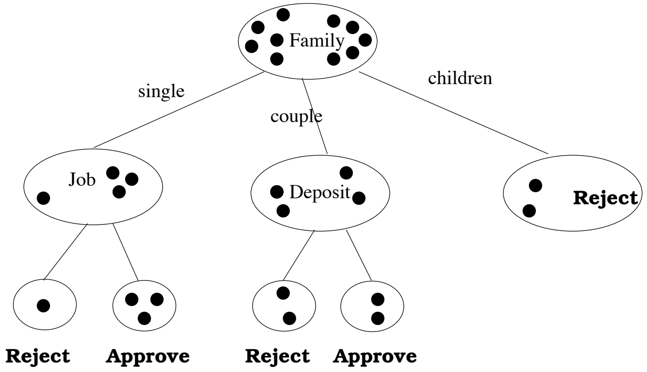
Learning Decision Trees

- Job:
  - true: 6/10 instances (4 approved, 2 rejected)
  - false: 4/10 instances (1 approved, 3 rejected)
  - impurity of node 1: 2/6 × 4/6
  - impurity of node 2: 3/4 × 1/4
  - weighted impurity: 6/10 × (2/6 × 4/6) + 4/10 × (3/4 × 1/4) = 21%
- Deposit
  - weighted impurity: 3/10 × (1/3 × 2/3) + 7/10 × (4/7 × 3/7) = 24%
- Family

  – weighted impurity: 4/10 × (1/4 × 3/4) + 4/10 × (2/4 × 2/4) + 2/10 × (0/2 × 2/2) = 18%

- Which one should we choose?

Decision Tree Building



- Identify and label pure nodes
- Recurse on impure nodes
  - −→ Consider attributes "Job" and "Deposit"

Summary

- Decision tree learning vs learned decision trees
- Method of building a decision tree: DT learning algorithm
- Purity measures: weighted average impurities

- Next Lecture:
  - numeric attributes: splitting points
  - Perceptron learning

- Suggested reading: section 20.5 (2nd edition) or section 18.7 (3rd edition) and web