Data Mining and Machine Learning

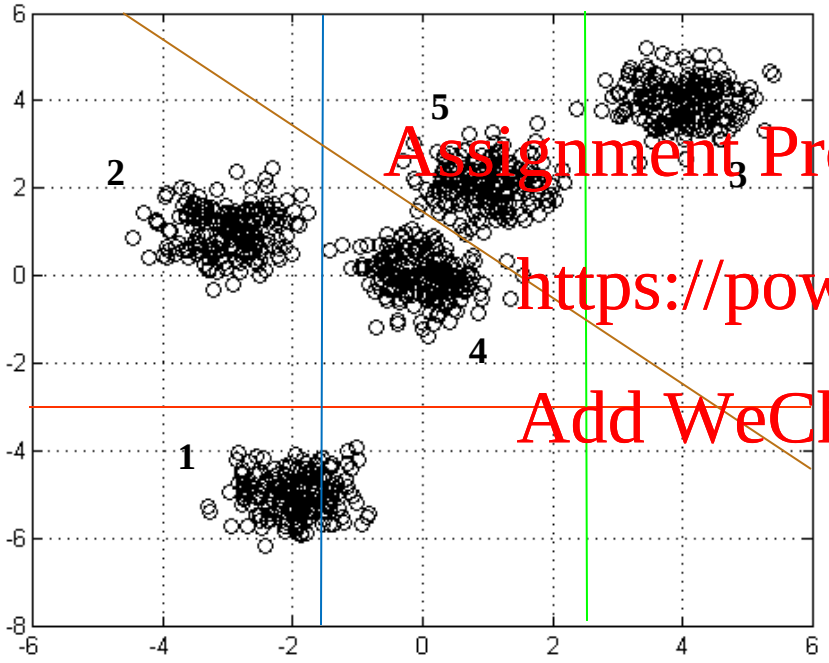
Decision trees

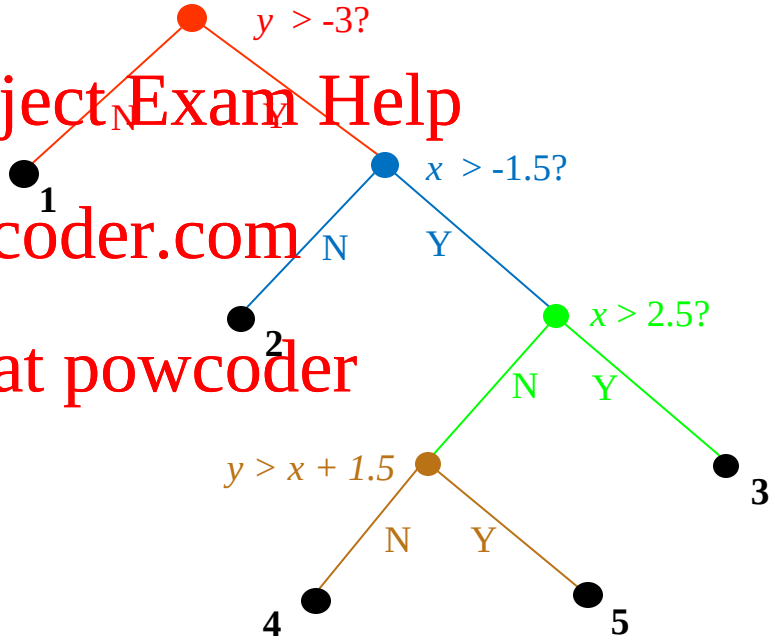
Peter Jančovič

Outline of lecture

- Introduction to Decision Trees (DTs)
  - A third approach to partitioning data
  - Basic principles
  - Decision tree issues
    - Types of question
    - Automatic construction of DTs from data
  - Example from Speech Recognition: Phone Decision Trees

A simple binary decision tree





Binary decision trees

- A binary decision tree consists of:
  - A binary tree (non-terminal nodes have 2 ‘children’)
  - For each node s:
    - A “yes/no” question q_s which can be applied to any data sample
    - A rule that determines which link is traversed if the answer to q_s is “YES” (then the other link is traversed if the answer is “NO”)
- Questions need not be numerical

Binary decision trees

- The decision tree on the previous slide is a simple 'hand-crafted' example, but what if the data was in a much higher dimensional vector space?
- How do we decide the best question to associate with each node of the tree?
- How do we know when the tree is complete?
- What are the alternatives to numerical questions?

Example from speech recognition

- In speech recognition a word is modelled as a sequence of phones: e.g. “six” = /s/ /I/ /k/ /s/
- Each phone is modelled as a statistical Hidden Markov Model (HMM). Each model has many parameters which must be estimated from data.
- In English there are approximately 45 phones
- Fundamental problem is “co-articulation” - the acoustic signal corresponding to a particular phone depends on preceding and following phones

Example (continued)

- Positions of articulators when a phone is produced depend on positions before and after it is produced
- Context-sensitive phone modelling
- Triphones – separate phone model for each preceding and following phone (so the /I/ in s I k s would be a model of /I/ preceded by /s/ and followed by /k/
- Potentially huge model set (45^3 = 91,125 models)
- Phone decision trees combine contexts that result in similar coarticulation

Phone decision trees

- Collect very large set of acoustic patterns, each corresponding to a phone in context (triphones)
  - "Acoustic pattern" means a sequence of feature vectors that capture the salient short-term properties of the speech signal
  - Vectors are high-dimensional – typically ~40D
- This is a "big data" problem. Identifying contexts that induce similar coarticulation is a clustering problem.

Decision tree "questions"

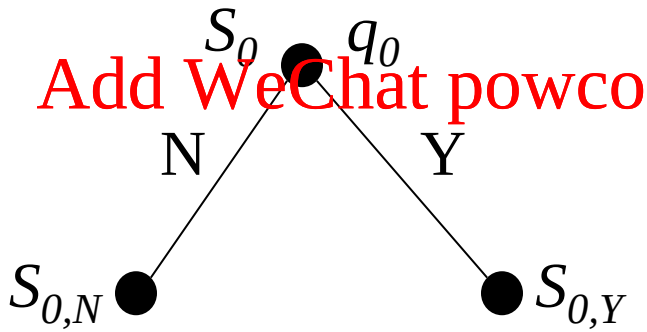
- Each node associated with a "question"
- Knowledge from phonetics and linguistics defines a set of potential decision-tree node "questions"
- For example:
  - Q: "Is the left context one of {/p/, /b/, /t/, /d/}?"
  - Q: "Is the right-context {/m/}?"
- Call these questions q_1,..., q_N

Phone decision tree construction

- Associate complete set of phones-in-context, S_0, with root node
- Each question q_n partitions S_0 into two subsets:
  - The set S_{0nY} of samples for which q_n is true, and
  - The set S_{0nN} of samples for which q_n is false
- Intuitively, if q_n is a "good question" then:
  - Samples in S_{0nY} will be similar to each other, and
  - Samples in S_{0nN} will be similar to each other
- Let $P_{0n} = P_{0nY} + P_{0nN}$ where P_{0nA} measures similarity between samples in S_{0nA} ($A = Y$ or N)

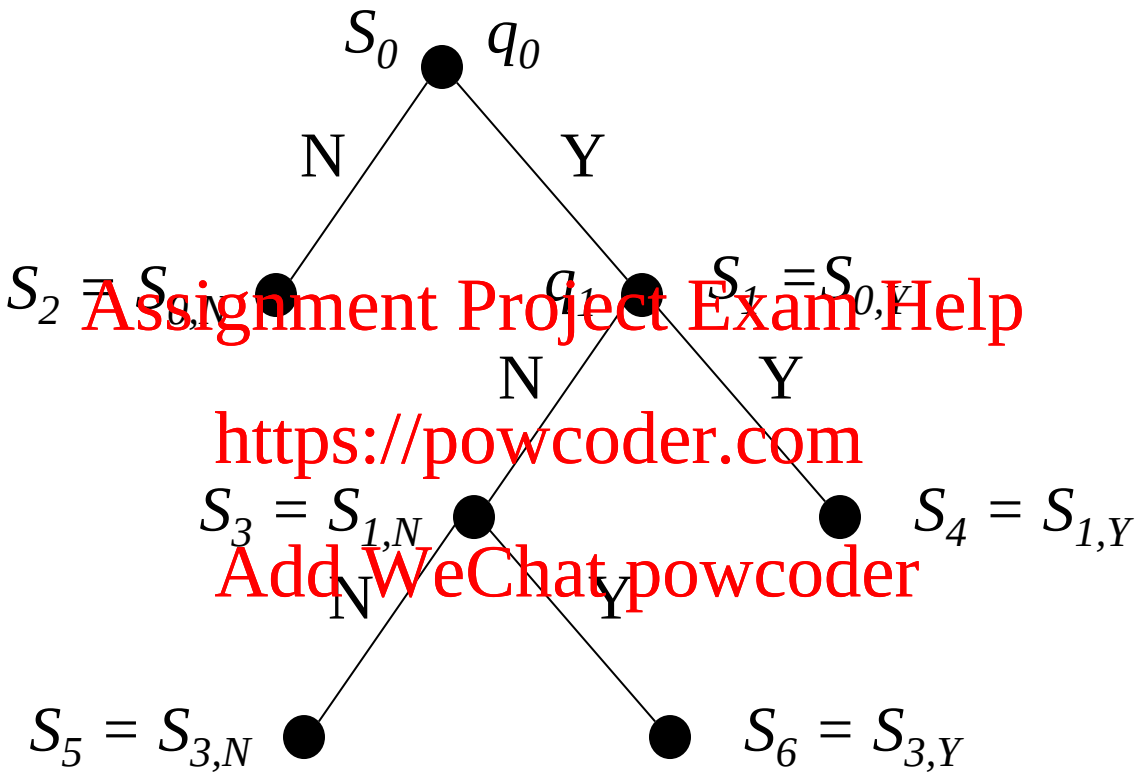
Phone decision tree construction

- P_{0nY} could be a measure of how well S_{0nY} is modelled by a single Gaussian PDF
- Calculate P_{0n} for every question q_n
- Let q_0 be the question q_n for which P_{0n} is maximised



- Repeat whole process for $S_{0,N}$ and $S_{0,Y}$, and so on

Phone decision tree construction



- Stop if too few samples in S_n, or benefit of splitting S_n is too small

Phone decision trees

- Separate HMM built for each terminal (leaf) node
- Given a new phone-in-context, negotiate the tree according to the answers to the questions until a terminal node is reached – the corresponding model is the one to use for this phone-in-context
- Phone decision trees illustrate two important points:
  - Example of DT with non-numerical questions
  - Illustration of how DT can be constructed automatically from data

Summary

- Introduction to Decision Trees
- Decision Tree issues
- Example: Phone decision Trees