Decision Trees
(Part I)
Q: How will I earn the 5% Participation points?

A: Good question! One way is by filling out the required poll on what WIFI enabled devices you have on Piazza. [https://piazza.com/class/jqnu4yso196rm?cid=15](https://piazza.com/class/jqnu4yso196rm?cid=15)

Other points will be earned through in-class polls, some “grace days”, and other opportunities to gain participation points.

Starting next week, please come to class with a WIFI enabled smartphone or tablet. We’ll announce on Piazza what to do if you don’t have such a device.
Reminders

• Homework 1: Background
  – Out: Wed, Jan 16 (2nd lecture)
  – Due: Wed, Jan 23 at 11:59pm
  – Two parts:
    1. written part to Gradescope,
    2. programming part to Autolab
  – unique policy for this assignment:
    1. two submissions for written (see writeup for details)
    2. unlimited submissions for programming (i.e. keep submitting until you get 100%),
  – unique policy for this assignment: we will grant (essentially) any and all extension requests
Big Ideas

1. How to formalize a learning problem
2. How to learn an expert system (i.e. Decision Tree)
3. Importance of inductive bias for generalization
4. Overfitting
FUNCTION APPROXIMATION
Function Approximation

Quiz: Implement a simple function which returns $\sin(x)$.

A few constraints are imposed:
1. You can’t call any other trigonometric functions
2. You can call an existing implementation of $\sin(x)$ a few times (e.g. 100) to test your solution
3. You only need to evaluate it for $x$ in $[0, 2\pi]$
Medical Diagnosis

- Setting:
  - Doctor must decide whether or not to prescribe a treatment
  - Looks at attributes of a patient to make a medical diagnosis
  - Prescribes treatment if diagnosis is positive
- Key problem area for Machine Learning
- Potential to reshape health care
ML as Function Approximation

Chalkboard

– ML as Function Approximation
  • Problem setting
  • Input space
  • Output space
  • Unknown target function
  • Hypothesis space
  • Training examples
DECISION TREES
Decision Trees

**Chalkboard**

- Example: Medical Diagnosis
- Does memorization = learning?
- Decision Tree as a hypothesis
- Function approximation for DTs
Learned from medical records of 1000 women
Negative examples are C-sections

[833+,167-] .83+ .17-
Fetal_Presentation = 1: [822+,116-] .88+ .12-
  | Previous_Csection = 0: [767+,81-] .90+ .10-
  | | Primiparous = 0: [399+,13-] .97+ .03-
  | | Primiparous = 1: [368+,68-] .84+ .16-
  | | | Fetal_Distress = 0: [334+,47-] .88+ .12-
  | | | Birth_Weight < 3349: [201+,10.6-] .95+ .05-
  | | | Birth_Weight >= 3349: [133+,36.4-] .78+ .22-
  | | | | Fetal_Distress = 1: [34+,21-] .62+ .38-
  | | Previous_Csection = 1: [55+,35-] .61+ .39-
Fetal_Presentation = 2: [3+,29-] .11+ .89-
Fetal_Presentation = 3: [8+,22-] .27+ .73-

(Sims et al., 2000)