10-601 Introduction to Machine Learning

Machine Learning Department
School of Computer Science
Carnegie Mellon University

Decision Trees (Part II)

Matt Gormley
Lecture 3
Jan. 22, 2020

Q&A

Q: In our medical diagnosis example, suppose two of our doctors (i.e. experts) disagree about whether (+) or not (-) the patient is sick. How would the decision tree represent this situation?

A: Today we will define decision trees that predict a single class by a majority vote at the leaf. More generally, the leaf could provide a probability distribution over output classes $p(y|\mathbf{x})$

Q&A

Q: How do these In-Class Polls work?

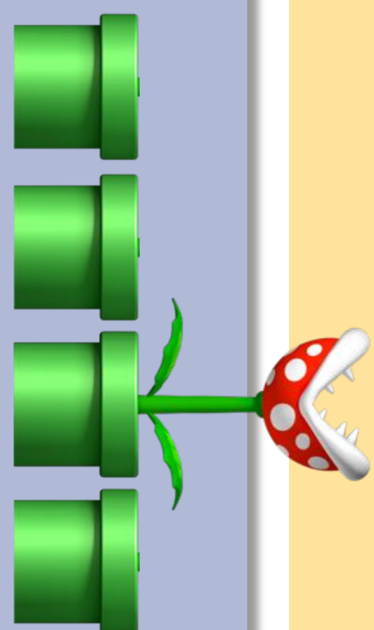
A:
- Sign into **Google Form** (link from Schedule page) using **Andrew Email (http://p3.mlcourse.org)**
- Answer **during lecture for full credit**, or the same day (i.e. before 11:59pm) for half credit
- Avoid the **calamity option** which gives negative points!
- 8 "free poll points" but can't use more than 3 free polls consecutively
- Submit a **poll card** if and only if you do not have a smartphone/tablet

First In-Class Poll

Question:

Which of the following did you bring to class today?

A. Smartphone
B. Flip phone
C. Pay phone
D. No phone



Answer:

Reminders

- **Homework 1: Background**
  - **Out: Wed, Jan 15 (2nd lecture)**
  - **Due: Wed, Jan 22 at 11:59pm**
  - unique policy for this assignment: we will grant (essentially) any and all extension requests
- **Homework 2: Decision Trees**
  - **Out: Wed, Jan. 22**
  - **Due: Wed, Feb. 05 at 11:59pm**

DECISION TREES

Decision Trees

Chalkboard

- Decision Tree as a hypothesis
- Function approximation for DTs

Tree to Predict C-Section Risk

Learned from medical records of 1000 women (Sims et al., 2000)

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+ .
| | | | Birth_Weight >= 3349: [133+,36.4-] .78+
| | | 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-
```

Figure from Tom Mitchell

Decision Trees

Chalkboard

– Decision Tree Learning

Decision Tree Learning Example

Dataset:

Output Y, Attributes A, B, C

Y	A	B	C
-	1	0	0
-	1	0	1
-	1	0	0
+	0	0	1
+	1	1	0
+	1	1	1
+	1	1	0
+	1	1	1

In-Class Exercise

Using **error rate** as the splitting criterion, what decision tree would be learned?

SPLITTING CRITERIA FOR DECISION TREES

Decision Tree Learning

- *Definition*: a **splitting criterion** is a function that measures the effectiveness of splitting on a particular attribute
- Our decision tree learner **selects the "best" attribute** as the one that maximizes the splitting criterion
- Lots of options for a splitting criterion:
  - error rate (or *accuracy* if we want to pick the tree that *maximizes* the criterion)
  - Gini gain
  - Mutual information
  - random
  - …

Decision Tree Learning Example

Dataset:

Output Y, Attributes A and B

Y	A	B
-	1	0
-	1	0
+	1	0
+	1	0
+	1	1
+	1	1
+	1	1
+	1	1

In-Class Exercise

Which attribute would **error rate** select for the next split?

1. A
2. B
3. A or B (tie)
4. Neither

Decision Tree Learning Example

Dataset:

Output Y, Attributes A and B

Y	A	B
-	1	0
-	1	0
+	1	0
+	1	0
+	1	1
+	1	1
+	1	1
+	1	1

Gini Impurity

Chalkboard

- Expected Misclassification Rate:
  - Predicting a Weighted Coin with another Weighted Coin
  - Predicting a Weighted Dice Roll with another Weighted Dice Roll
- Gini Impurity
- Gini Impurity of a Bernoulli random variable
- Gini Gain as a splitting criterion

Decision Tree Learning Example

Dataset:

Output Y, Attributes A and B

Y	A	B
-	1	0
-	1	0
+	1	0
+	1	0
+	1	1
+	1	1
+	1	1
+	1	1

In-Class Exercise

Which attribute would **Gini gain** select for the next split?

1. A
2. B
3. A or B (tie)
4. Neither

Decision Tree Learning Example

Dataset:

Output Y, Attributes A and B

Y	A	B
-	1	0
-	1	0
+	1	0
+	1	0
+	1	1
+	1	1
+	1	1
+	1	1