CAP 5610
Assignment #1 Solution

January 23, 2025

Arman Sayan

---

1 Decision Tree Basics   [30 points]

The goal of this assignment is to test and reinforce your understanding of Decision Tree Classifiers.

1. [5 points] How many unique, perfect binary trees of depth 3 can be drawn if we have 5 attributes? By depth, we mean the depth of the splits, not including the nodes that only contain a label (see Figure 1). So, a tree that checks just one attribute is a depth one tree. By perfect binary tree, we mean every node has either 0 or 2 children, and every leaf is at the same depth. Note also that a tree with the same attributes but organized at different depths is considered "unique". Do not include trees that test the same attribute along the same path in the tree.
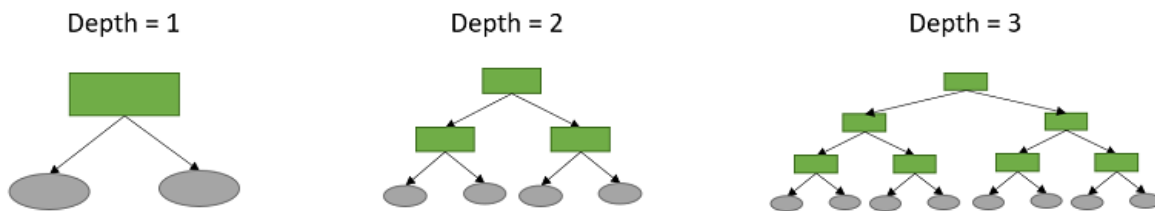

Figure 1: Example of perfect binary trees with different depths.

Ans:

2. [5 points] In general, for a problem with A attributes, how many unique perfect D depth trees can be drawn? Assume $A >> D$.

> **Ans:**

3. [10 points] Consider the following dataset from Table 1 for this problem. Given the five attributes on the left, we want to predict if the student got an A in the course. Create 2 decision trees for this dataset. For the first, only go to depth 1. For the second go to depth 2. For all trees, use the ID3 entropy algorithm from class. For each node of the tree, show the decision, the number of positive and negative examples and show the entropy at that node.

   Hint: There are a lot of calculations here. You may want to do this programatically.

Early	Finished HMK	Senior	Likes Coffee	Liked The Last Jedi	A
1	1	0	0	1	1
1	1	1	0	1	1
0	0	1	0	0	0
0	1	1	0	1	0
0	1	1	0	0	1
0	0	1	1	1	1
1	0	0	0	1	0
0	1	0	1	1	1
0	0	1	0	1	1
1	0	0	0	0	0
1	1	1	0	0	1
0	1	1	1	1	0
0	0	0	0	1	0
1	0	0	1	0	1

Table 1: Toy Data-set for Task 1: Decision Tree Basics.

> **Ans:**

4. [5 points] Make one more decision tree. Use the same procedure as in (b), but make it depth 3. Now, given these three trees, which would you prefer if you wanted to predict the grades of 10 new students who are not included in this data-set? Justify your choice.

> **Ans:**

5. [5 points] Consider a new definition of a "realizable" case: "For some fixed concept class C, such as decision trees, a realizable case is one where the algorithm gets a sample consistent with some concept $c \in C$. In other words, for decision trees, a case is realizable if there is some tree that perfectly classifies the data-set.

Ans:

2   Application of Decision Tree on Real-Word Data-set   [25 points]

1. [10 points] Train a decision tree classifier using the data file. Vary the cut-off depth from 2 to 10 and report the training accuracy for each cut-off depth k. Based on your results, select an optimal k.

> **Ans:** Please check the source code included in the .zip file named as
> **CAP_5610_Assignment_1_Solution_Arman_Sayan.ipynb**
> for the solution.

2. [8 points] Using the trained classifier with optimal cut-off depth k, classify the 99,762 instances from the test file and report the testing accuracy (the portion of testing instances classified correctly).

> **Ans:** Please check the source code included in the .zip file named as
> **CAP_5610_Assignment_1_Solution_Arman_Sayan.ipynb**
> for the solution.

3. [7 points] Do you see any over-fitting issues for this experiment? Report your observations.

> **Ans:** Please check the source code included in the .zip file named as
> **CAP_5610_Assignment_1_Solution_Arman_Sayan.ipynb**
> for the solution.

3   Independent Events and Bayes Theorem   [20 points]

1. [5 points] For events A, B prove:
$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}$$
($\neg A$ denotes the event that A does not occur.)

Ans:

We need to prove Bayes' theorem in the expanded form. We know from Bayes' rule that the definition of conditional probability is
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
By a simple induction, we can write $P(A \cap B)$ as
$$P(A \cap B) = P(B|A)P(A)$$
Furthermore, using the law of total probability, we can express $P(B)$ as
$$P(B) = P(A \cap B) + P(\neg A \cap B)$$
where by using the above induction again, we reveal that
$$P(\neg A \cap B) = P(B|\neg A)P(\neg A)$$
Using these definitions and substituting them in the original theorem, we can prove that
$$\boldsymbol{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$
$$\boldsymbol{P(A|B) = \frac{P(B|A)P(A)}{P(A \cap B) + P(\neg A \cap B)}}$$
$$\boldsymbol{P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}}$$

---

2. Let X, Y , and Z be random variables taking values in $0, 1$. The following table lists the probability of each possible assignment of 0 and 1 to the variables X, Y, and Z:

	$Z = 0$		$Z = 1$	
	$X = 0$	$X = 1$	$X = 0$	$X = 1$
$Y = 0$	0.1	0.05	0.1	0.1
$Y = 1$	0.2	0.1	0.175	0.175

i. [5 points] Is X independent of Y? Why or why not?

Ans: To determine if X and Y are independent, we need to check whether the joint probability $P(X = x, Y = y)$ equals the product of the marginal probabilities $P(X = x) \cdot P(Y = y)$ for all possible values of x and y. If this holds true for all combinations, X and Y are independent. Otherwise, they are dependent.

As the first step, we need to calculate the marginal probabilities. The formula for marginal probability $P(X = x, Y = y)$ is given as

$$P(X = x, Y = y) = P(X = x, Y = y, Z = 0) + P(X = x, Y = y, Z = 1)$$

For each combination of X and Y, we can calculate the marginal probabilities as follows:

- $P(X = 0, Y = 0) = P(X = 0, Y = 0, Z = 0) + P(X = 0, Y = 0, Z = 1) = 0.1 + 0.1 = 0.2$

- $P(X = 1, Y = 0) = P(X = 1, Y = 0, Z = 0) + P(X = 1, Y = 0, Z = 1) = 0.05 + 0.1 = 0.15$

- $P(X = 0, Y = 1) = P(X = 0, Y = 1, Z = 0) + P(X = 0, Y = 1, Z = 1) = 0.2 + 0.175 = 0.375$

- $P(X = 1, Y = 1) = P(X = 1, Y = 1, Z = 0) + P(X = 1, Y = 1, Z = 1) = 0.1 + 0.175 = 0.275$

Now, we can calculate the marginal probabilities $P(X)$ and $P(Y)$ as follows:

- $P(X = 0) = P(X = 0, Y = 0) + P(X = 0, Y = 1) = 0.2 + 0.375 = 0.575$

- $P(X = 1) = P(X = 1, Y = 0) + P(X = 1, Y = 1) = 0.15 + 0.275 = 0.425$

- $P(Y = 0) = P(X = 0, Y = 0) + P(X = 1, Y = 0) = 0.2 + 0.15 = 0.35$

- $P(Y = 1) = P(X = 0, Y = 1) + P(X = 1, Y = 1) = 0.375 + 0.275 = 0.65$

The second step is to check the independce, which means for X and Y to be independent, the following must hold:
$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$
Check $P(X = 0, Y = 0)$:
$$P(X = 0, Y = 0) = 0.2$$
$$P(X = 0) \cdot P(Y = 0) = 0.575 \cdot 0.35 = 0.20125$$
$$\boldsymbol{P(X = 0, Y = 0) \neq P(X = 0) \cdot P(Y = 0)}$$
Check $P(X = 1, Y = 0)$:
$$P(X = 1, Y = 0) = 0.15$$
$$P(X = 1) \cdot P(Y = 0) = 0.425 \cdot 0.35 = 0.14875$$
$$\boldsymbol{P(X = 1, Y = 0) \neq P(X = 1) \cdot P(Y = 0)}$$
Check $P(X = 0, Y = 1)$:
$$P(X = 0, Y = 1) = 0.375$$
$$P(X = 0) \cdot P(Y = 1) = 0.575 \cdot 0.65 = 0.37375$$
$$\boldsymbol{P(X = 0, Y = 1) \neq P(X = 0) \cdot P(Y = 1)}$$
Check $P(X = 1, Y = 1)$:
$$P(X = 1, Y = 1) = 0.275$$
$$P(X = 1) \cdot P(Y = 1) = 0.425 \cdot 0.65 = 0.27625$$
$$\boldsymbol{P(X = 1, Y = 1) \neq P(X = 1) \cdot P(Y = 1)}$$
The values $P(X = x, Y = y)$ and $P(X = x)P(Y = y)$ are not exactly equal for all combinations. Therefore, **X and Y are not independent**.

ii. [5 points] Is X conditionally independent of Y given Z? Why or why not?

Ans: To determine whether X is conditionally independent of Y given Z, we need to verify if the following holds for all values of X, Y, and Z:
$$P(X|Y,Z) = P(X|Z)$$
The formula for conditional probability $P(X|Y,Z)$ is given as
$$P(X|Y,Z) = \frac{P(X,Y,Z)}{P(Y,Z)}$$
The first step is to calculate the joint probabilities $P(X,Y,Z)$ for all combinations of X, Y, and Z. We can use the values from the table as it is.

The second step is to calculate the marginal probabilities $P(Y,Z)$ for all combinations of Y and Z.

- $P(Y = 0, Z = 0) = P(X = 0, Y = 0, Z = 0) + P(X = 1, Y = 0, Z = 0) = 0.1 + 0.05 = 0.15$

- $P(Y = 1, Z = 0) = P(X = 0, Y = 1, Z = 0) + P(X = 1, Y = 1, Z = 0) = 0.2 + 0.1 = 0.3$

- $P(Y = 0, Z = 1) = P(X = 0, Y = 0, Z = 1) + P(X = 1, Y = 0, Z = 1) = 0.1 + 0.1 = 0.2$

- $P(Y = 1, Z = 1) = P(X = 0, Y = 1, Z = 1) + P(X = 1, Y = 1, Z = 1) = 0.175 + 0.175 = 0.35$

The third step is to calculate the conditional probabilities $P(X|Y,Z)$ and $P(X|Z)$. We will have 2 cases to check.

The first case is when $Z = 0$:

First, we compute the probabilities $P(Y, Z = 0)$ which are obtained by summing over all X:

- $P(Y = 0, Z = 0) = P(X = 0, Y = 0, Z = 0) + P(X = 1, Y = 0, Z = 0) = 0.1 + 0.05 = 0.15$

- $P(Y = 1, Z = 0) = P(X = 0, Y = 1, Z = 0) + P(X = 1, Y = 1, Z = 0) = 0.2 + 0.1 = 0.3$

$$P(Z = 0) = P(Y = 0, Z = 0) + P(Y = 1, Z = 0) = 0.15 + 0.3 = 0.45$$

Second, we compute the probabilities $P(X, Z = 0)$ which are obtained by summing over all Y:

- $P(X = 0, Z = 0) = P(X = 0, Y = 0, Z = 0) + P(X = 0, Y = 1, Z = 0) = 0.1 + 0.2 = 0.3$

- $P(X = 1, Z = 0) = P(X = 1, Y = 0, Z = 0) + P(X = 1, Y = 1, Z = 0) = 0.05 + 0.1 = 0.15$

Third, we compute the probability $P(X|Z = 0)$:

- $P(X = 0|Z = 0) = \dfrac{P(X = 0, Z = 0)}{P(Z = 0)} = \dfrac{0.3}{0.45} = 0.666$

- $P(X = 1|Z = 0) = \dfrac{P(X = 1, Z = 0)}{P(Z = 0)} = \dfrac{0.15}{0.45} = 0.333$

Fourth, we compute the probability $P(X|Y, Z = 0)$:

- For $Y = 0$:

$$P(X = 0|Y = 0, Z = 0) = \frac{P(X = 0, Y = 0, Z = 0)}{P(Y = 0, Z = 0)} = \frac{0.1}{0.15} = 0.666$$

$$P(X = 1|Y = 0, Z = 0) = \frac{P(X = 1, Y = 0, Z = 0)}{P(Y = 0, Z = 0)} = \frac{0.05}{0.15} = 0.333$$

- For $Y = 1$:

$$P(X = 0|Y = 1, Z = 0) = \frac{P(X = 0, Y = 1, Z = 0)}{P(Y = 1, Z = 0)} = \frac{0.2}{0.3} = 0.666$$

$$P(X = 1|Y = 1, Z = 0) = \frac{P(X = 1, Y = 1, Z = 0)}{P(Y = 1, Z = 0)} = \frac{0.1}{0.3} = 0.333$$

Since $P(X|Y, Z = 0) = P(X|Z = 0)$ for all values of X, the condition holds for $Z = 0$.
The second case is when $Z = 1$:
First, we compute the probabilities $P(Y, Z = 1)$ which are obtained by summing over all X:

- $P(Y = 0, Z = 1) = P(X = 0, Y = 0, Z = 1) + P(X = 1, Y = 0, Z = 1) = 0.1 + 0.1 = 0.2$

- $P(Y = 1, Z = 1) = P(X = 0, Y = 1, Z = 1) + P(X = 1, Y = 1, Z = 1) = 0.175 + 0.175 = 0.35$

$$P(Z = 1) = P(Y = 0, Z = 1) + P(Y = 1, Z = 1) = 0.2 + 0.35 = 0.55$$

Second, we compute the probabilities $P(X, Z = 1)$ which are obtained by summing over all Y:

- $P(X = 0, Z = 1) = P(X = 0, Y = 0, Z = 1) + P(X = 0, Y = 1, Z = 1) = 0.1 + 0.175 = 0.275$

- $P(X = 1, Z = 1) = P(X = 1, Y = 0, Z = 1) + P(X = 1, Y = 1, Z = 1) = 0.1 + 0.175 = 0.275$

Third, we compute the probability $P(X|Z = 1)$:

- $P(X = 0|Z = 1) = \dfrac{P(X = 0, Z = 1)}{P(Z = 1)} = \dfrac{0.275}{0.55} = 0.5$

- $P(X = 1|Z = 1) = \dfrac{P(X = 1, Z = 1)}{P(Z = 1)} = \dfrac{0.275}{0.55} = 0.5$

Fourth, we compute the probability $P(X|Y, Z = 1)$:

- For $Y = 0$:

$$P(X = 0|Y = 0, Z = 1) = \frac{P(X = 0, Y = 0, Z = 1)}{P(Y = 0, Z = 1)} = \frac{0.1}{0.2} = 0.5$$

$$P(X = 1|Y = 0, Z = 1) = \frac{P(X = 1, Y = 0, Z = 1)}{P(Y = 0, Z = 1)} = \frac{0.1}{0.2} = 0.5$$

- For $Y = 1$:

$$P(X = 0|Y = 1, Z = 1) = \frac{P(X = 0, Y = 1, Z = 1)}{P(Y = 1, Z = 1)} = \frac{0.175}{0.35} = 0.5$$

$$P(X = 1|Y = 1, Z = 1) = \frac{P(X = 1, Y = 1, Z = 1)}{P(Y = 1, Z = 1)} = \frac{0.175}{0.35} = 0.5$$

Like in the first case, Since $P(X|Y, Z = 1) = P(X|Z = 1)$ for all values of X, the condition holds for $Z = 1$.
For all values of Z, we proved that
$$P(X|Y, Z) = P(X|Z)$$
Thus, **X is conditionally independent of Y given Z**.

iii. [5 points] Calculate $P(X \neq Y | Z = 0)$.

Ans:
To compute $P(X \neq Y | Z = 0)$, we need to calculate the probability of events where $X \neq Y$, such as $X = 0$ and $Y = 1$, or $X = 1$ and $Y = 0$, conditioned on $Z = 0$.

We can calculate the probability as follows:
$$P(X \neq Y | Z = 0) = \frac{P(X \neq Y, Z = 0)}{P(Z = 0)}$$
Furthermore, we can calculate $P(X \neq Y, Z = 0)$ as
$$P(X \neq Y, Z = 0) = P(X = 0, Y = 1, Z = 0) + P(X = 1, Y = 0, Z = 0)$$
From the table, we can find the values as

- $P(X = 0, Y = 1, Z = 0) = 0.2$

- $P(X = 1, Y = 0, Z = 0) = 0.05$

$$P(X \neq Y, Z = 0) = 0.2 + 0.05 = 0.25$$

In addition, we need to calculate $P(Z = 0)$ by summing all joint probabilities where $Z = 0$:
$$P(Z = 0) = P(X = 0, Y = 0, Z = 0) + P(X = 1, Y = 0, Z = 0)$$
$$+ P(X = 0, Y = 1, Z = 0) + P(X = 1, Y = 1, Z = 0)$$
$$= 0.1 + 0.05 + 0.2 + 0.1 = 0.45$$
Finally, by substituting the values, we can calculate $P(X \neq Y | Z = 0)$ as
$$\boldsymbol{P(X \neq Y | Z = 0) = \frac{0.25}{0.45} \approx 0.5556}$$

4   Implementing Naive Bayes   [25 points]

You will now learn how to use Naive Bayes Algorithm to solve a real-world problem: text cat- egorization. Text categorization (also referred to as text classification) i s t he t ask o f assigning documents to one or more topics. For our homework, we will use a benchmark dataset that is fre- quently used in text categorization problems. This dataset, Reuters-21578, consists of documents that appeared in Reuters newswire in 1987. Each document was then manually categorized into a topic among over 100 topics. In this homework, we are only interested in earn and acquisition (acq) topics, so we will use a shortened version of the dataset (documents assigned to topics other than "earn" or "acq" are not in the dataset provided for the homework). As features, we will use the frequency (counts) of each word that occurred in the document. This model is known as the bag-of-words model and it is frequently used in text categorization. You can download Assignment 2 data from the Canvas. In this folder, you will find:

- **train.csv:** Training data. Each row represents a document, and each column separated by commas represents features (word counts). There are 4527 documents and 5180 words.

- **train labels.txt:** labels for the training data

- **test.csv:** Test data, 1806 documents and 5180 words

Implement Naive Bayes Algorithm. Train your classifier on the training set that is given and report training accuracy, testing accuracy, and the amount of time spent training the classifier.

Ans: Please check the source code included in the .zip file named as

CAP_5610_Assignment_1_Solution_Arman_Sayan.ipynb

for the solution.