CSCE 421: Machine Learning (Spring 2025)
Homework #1

Sutton Elliott

Due: February 4, 2025

Instructions

- Submit a PDF report to Canvas named `HW#_FirstName_LastName.pdf`.
- Ensure the PDF is directly uploaded (do not zip the file).
- Use Word or LaTeX to typeset your report for clarity. Hand-written reports must be clearly scanned.
- Multiple submissions are allowed; only the latest submission will be graded.
- Start submissions at least 15–30 minutes before the deadline to avoid issues. Late submissions via email are not accepted.

Questions

1. [10 pts]

Suppose that we have three colored boxes r (red), b (blue), and g (green). Box r contains 3 apples, 4 oranges, and 3 limes, box b contains 1 apple, 1 orange, and 0 limes, and box g contains 3 apples, 3 oranges, and 4 limes. If a box is chosen at random with probabilities $p(r) = 0.2, p(b) = 0.2, p(g) = 0.6$, and a piece of fruit is removed from the box (with equal probability of selecting any of the items in the box), then what is the probability of selecting an apple? If we observe that the selected fruit is in fact an orange, what is the probability is came from the green box?

Using the law of total probability,

$$P(A) = P(A|r)P(r) + P(A|b)P(b) + P(A|g)P(g), \tag{1}$$

Also,

$$P(A|r) = \frac{3}{10}, \quad P(A|b) = \frac{1}{2}, \quad P(A|g) = \frac{3}{10}.$$

Thus,

$$\begin{aligned} P(A) &= \left(\frac{3}{10} \times 0.2\right) + \left(\frac{1}{2} \times 0.2\right) + \left(\frac{3}{10} \times 0.6\right) \\ &= 0.06 + 0.1 + 0.18 = \boxed{0.34.} \end{aligned}$$

Probability that an Orange came from Box g
Using Bayes' theorem,

$$P(g|O) = \frac{P(O|g)P(g)}{P(O)}. \tag{2}$$

First, we compute $P(O)$:

$$P(O) = P(O|r)P(r) + P(O|b)P(b) + P(O|g)P(g),$$

where:

$$P(O|r) = \frac{4}{10} = 0.4, \quad P(O|b) = \frac{1}{2} = 0.5, \quad P(O|g) = \frac{3}{10} = 0.3.$$

Thus,

$$\begin{aligned} P(O) &= (0.4 \times 0.2) + (0.5 \times 0.2) + (0.3 \times 0.6) \\ &= 0.08 + 0.1 + 0.18 = 0.36. \end{aligned}$$

Now, compute $P(g|O)$:

$$P(g|O) = \frac{(0.3 \times 0.6)}{0.36} = \frac{0.18}{0.36} = \boxed{0.5}.$$

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

2. [20 pts]

Given the following data set containing three attributes and one class, use the naïve Bayes classifier to determine the class (Yes/No) of **Stolen** for a **Red Domestic SUV**.

We use the Naïve Bayes classifier to compute the probability of the class being **Yes** or **No** for a Red Domestic SUV.

Step 1: Compute Prior Probabilities Total examples: 10

$$P(\text{Yes}) = \frac{5}{10} = 0.5,$$
$$P(\text{No}) = \frac{5}{10} = 0.5.$$

Step 2: Compute Likelihood Probabilities

(a) Probability of Red given Yes and No:

$$P(\text{Red}|\text{Yes}) = \frac{3}{5} = 0.6,$$
$$P(\text{Red}|\text{No}) = \frac{2}{5} = 0.4.$$

(b) Probability of Domestic given Yes and No:

$$P(\text{Domestic}|\text{Yes}) = \frac{2}{5} = 0.4,$$
$$P(\text{Domestic}|\text{No}) = \frac{3}{5} = 0.6.$$

(c) Probability of SUV given Yes and No:

$$P(\text{SUV}|\text{Yes}) = \frac{1}{5} = 0.2,$$
$$P(\text{SUV}|\text{No}) = \frac{2}{5} = 0.4.$$

Step 3: Compute Posterior Probabilities Using Bayes' theorem:

$$\begin{aligned} P(\text{Yes}|\text{Red, Domestic, SUV}) &\propto P(\text{Yes})P(\text{Red}|\text{Yes})P(\text{Domestic}|\text{Yes})P(\text{SUV}|\text{Yes}) \\ &= (0.5)(0.6)(0.4)(0.2) = 0.024. \end{aligned}$$

$$\begin{aligned} P(\text{No}|\text{Red, Domestic, SUV}) &\propto P(\text{No})P(\text{Red}|\text{No})P(\text{Domestic}|\text{No})P(\text{SUV}|\text{No}) \\ &= (0.5)(0.4)(0.6)(0.4) = 0.072. \end{aligned}$$

Step 4: Conclusion Since $P(\text{No}|\text{Red, Domestic, SUV}) > P(\text{Yes}|\text{Red, Domestic, SUV})$, the Naïve Bayes classifier predicts:

$$\textbf{No}$$

Thus, a Red Domestic SUV is classified as **Not Stolen**.

3. [20 pts]

This question is about naïve Bayes classifier. Please do the following:

(a) State the simplifying assumption made by the naïve Bayes classifier.

The simplifying assumption made by the Naïve Bayes classifier is that all features are conditionally independent given the class label.

$$P(X_1, X_2, \ldots, X_d|C) = \prod_{i=1}^{d} P(X_i|C).$$

This assumption significantly reduces the complexity of computing the joint probability distribution.

(b) Given a binary-class classification problem where the class labels are binary, the dimension of features is d, and each attribute can take k different values, provide the number of parameters to be estimated **with** and **without** the simplifying assumption. Briefly justify why the simplifying assumption is necessary.

In a binary-class classification problem where each of the d attributes can take k different values, the number of parameters needed to estimate the full joint probability distribution is:

$$(k^d - 1) \times 2.$$

This accounts for the probability distribution of each feature combination given each class, minus one due to the sum-to-one constraint.

Using the Naïve Bayes assumption, we estimate the probability of each attribute independently given the class. The number of parameters required is:

$$(d \times (k - 1)) \times 2.$$

Here, we estimate d conditional probability distributions (one for each feature), each with $k - 1$ parameters (since each distribution sums to one), for the two class labels.

Justification for the Simplifying Assumption: The assumption is necessary because estimating the full joint distribution requires an exponential number of parameters, making large datasets more difficult. The Naïve Bayes assumption reduces this to a manageable number, making the model more efficient and easy to train.

Assuming that the probability of each evidence word is independent of other word occurrences given the category of the text, compute the (posterior) probability for each of the possible categories each of the following short texts; and based on that, their most likely classification. Assume that the categories are disjoint and exhaustive (i.e., every text is either physics, or biology or chemistry and no text can be more than one). Assume that words are first stemmed to reduce them to their base form (atoms→atom) and ignore any words that are not in the table:

c	Physics	Biology	Chemistry
P(c)	0.35	0.40	0.25
P(atom | c)	0.1	0.01	0.2
P(carbon | c)	0.005	0.03	0.05
P(proton | c)	0.05	0.001	0.05
P(life | c)	0.001	0.1	0.008
P(earth | c)	0.005	0.006	0.003

A: ***the carbon atom is the foundation of life on earth.***
B. the carbon atom contains 12 protons.

4. [25 pts]

Assume we want to classify science texts into three categories—physics, biology, and chemistry. The above probabilities have been estimated by analyzing a corpus of pre-classified web pages from Yahoo.

For each text, we extract the relevant words:
Text A: "The carbon atom is the foundation of life on earth."
Relevant words: carbon, atom, life, earth
Text B: "The carbon atom contains 12 protons."
Relevant words: carbon, atom, proton
We compute the posterior probabilities for each category using the Naive Bayes formula:

$$P(\text{Category} \mid \text{Text}) \propto P(\text{Category}) \cdot P(\text{word}_1 \mid \text{Category}) \cdot P(\text{word}_2 \mid \text{Category}) \cdot \cdots \cdot P(\text{word}_n \mid \text{Category})$$

For Text A:

$$P(\text{Physics} \mid A) \propto (0.35) \cdot (0.005) \cdot (0.1) \cdot (0.001) \cdot (0.005) = 8.75 \times 10^{-10}$$
$$P(\text{Biology} \mid A) \propto (0.40) \cdot (0.03) \cdot (0.01) \cdot (0.1) \cdot (0.006) = 7.2 \times 10^{-8}$$
$$P(\text{Chemistry} \mid A) \propto (0.25) \cdot (0.05) \cdot (0.2) \cdot (0.008) \cdot (0.003) = 6.0 \times 10^{-8}$$

Since $P(\text{Biology} \mid A) > P(\text{Chemistry} \mid A) > P(\text{Physics} \mid A)$, Text A is classified as **Biology**.
For Text B:

$$P(\text{Physics} \mid B) \propto (0.35) \cdot (0.005) \cdot (0.1) \cdot (0.005) = 8.75 \times 10^{-6}$$
$$P(\text{Biology} \mid B) \propto (0.40) \cdot (0.03) \cdot (0.01) \cdot (0.001) = 1.2 \times 10^{-7}$$
$$P(\text{Chemistry} \mid B) \propto (0.25) \cdot (0.05) \cdot (0.2) \cdot (0.05) = 1.25 \times 10^{-4}$$

Since $P(\text{Chemistry} \mid B) > P(\text{Physics} \mid B) > P(\text{Biology} \mid B)$, Text B is classified as **Chemistry**.

No.	Outlook	Temperature	Humidity	Windy	Play Golf?
1	sunny	hot	high	false	N
2	sunny	hot	high	true	N
3	overcast	hot	high	false	Y
4	rain	mild	high	false	Y
5	rain	cool	normal	false	Y
6	rain	cool	normal	true	N
7	overcast	cool	normal	true	Y
8	sunny	mild	high	false	N
9	sunny	cool	normal	false	Y
10	rain	mild	normal	false	Y
11	sunny	mild	normal	true	Y
12	overcast	mild	high	true	Y
13	overcast	hot	normal	false	Y
14	rain	mild	high	true	N

5. [25 pts]

Consider the following table of observations:

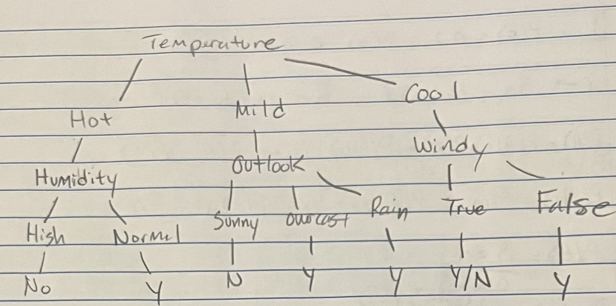
From the classified examples in the above table, construct two decision trees for the classification **"Play Golf"**:

(a) For the first tree, use **Temperature** as the root node (this is a bad choice). Continue constructing the tree as discussed in class, using information gain for subsequent nodes. Note: Different attributes can be used in different branches at a given level.

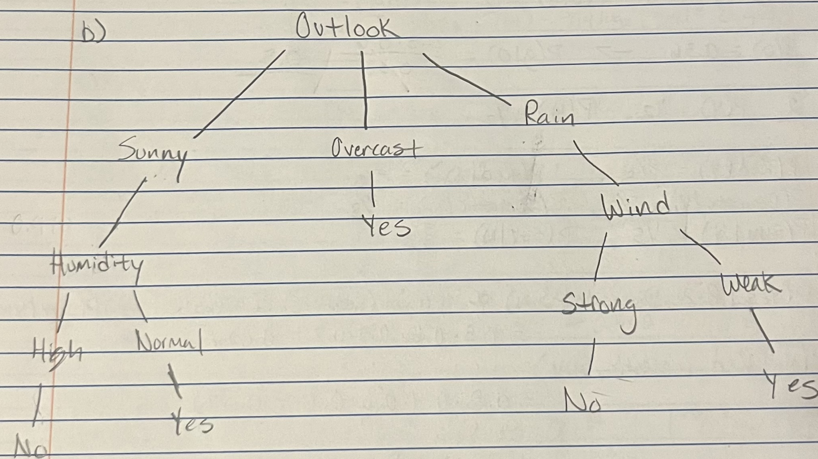
(b) For the second tree, use the Decision Tree Learning algorithm discussed in class. At each step, choose the attribute with the highest information gain. Compute the information gain by hand and draw the decision tree.

From the pictures below, to agree with the first question, temperature is a bad choice as the root of a decision tree. In the end, the tree ends up having a branch that does not conclude. The final branch for if windy it true can be yes or no so this is a bad decision tree simply for that reason.
The other tree is a much better tree using the entropy as a marker for which branch should be on which attribute. This ensures that each comes to a final decision. Calculations for that are below as well.

a)



b)



Sunny Branch:

Gain (Humid) = 0.970 → Humidity next for sunny

Gain (Temp) ≈ 0.570

Gain (Wind) = 0.019

5. Total Entropy: $-9/14 \log_2(9/14) - 5/14 \log_2(5/14) = 0.940$

Outlook:

2Y, 3N Sunny, 4Y, 0N Overcast, 3Y, 2N Rain

$E(\text{Outlook}) = 5/14\,(-2/5 \log_2(2/5) - 3/5 \log_2(3/5)) + 4/14\,(0) + 5/14\,(-3/5 \log_2(3/5)$
$- 2/5 \log_2(2/5) = 0.694 \rightarrow 0.940 - 0.694 = 0.246$

Temp: Hot, 2Y 2N, Mild, 4Y 2N, Cool, 3Y, 1N
$E(\text{Temp}): 4/14\,(-1/2 \log(1/2) - 1/2 \log_2(1/2)) + 6/14\,(-2/3 \log_2(2/3) - 1/3 \log_2(1/3))$
$+ 4/14\,(-3/4 \log_2(3/4) - 1/4 \log_2(1/4)) = 0.911 \rightarrow 0.940 - 0.911 = \underline{0.029}$

Humidity: High, 3Y4N, Normal 6Y 1N
$E(\text{Humd}) = 7/14\,(-3/7 \log_2(3/7) - 4/7 \log_2(4/7)) + 7/14\,(-6/7 \log_2(6/7) -$
$1/7 \log_2(1/7)) = 0.788 \rightarrow 0.940 - 0.788 = \underline{0.152}$

Wind: True 3Y 3N, False 6Y 2N
$E(\text{wind}) = 6/14\,(-3/6 \log_2(3/6) - 3/6 \log_2(3/6)) + 8/14\,(-6/8 \log_2(6/8)$
$- 2/8 \log(2/8)) = 0.892 \rightarrow 0.940 - 0.892 = \underline{0.048}$

Outlook as best root!

Rain Branch:
Gain(Wind) = 0.970 → Wind next for Rain
Gain(Temp) = 0.020
Gain(Humid) = 0.020

Since Overcast only leads to Yes then No calculations