# CSCE 421: Machine Learning
# Homework #1

Chayce J Leonard

Submitted: February 4, 2025

## Problem 1: Probability Calculations

### Part (a): Probability of Selecting an Apple

$$P(\text{Apple}) = \boxed{0.34}$$

### Part (b): Probability of Green Box Given Orange

$$P(g \mid \text{Orange}) = \boxed{0.5}$$

## Problem 2: Naïve Bayes Classification (20 pts)

### Solution

**Dataset Overview:**

- Total examples: 10
- Prior probabilities:

$$P(\text{Yes}) = 0.5, \quad P(\text{No}) = 0.5$$

**Likelihoods:**

$$P(\text{Red} \mid \text{Yes}) = 0.6, \quad P(\text{SUV} \mid \text{Yes}) = 0.2, \quad P(\text{Domestic} \mid \text{Yes}) = 0.4$$

$$P(\text{Red} \mid \text{No}) = 0.4, \quad P(\text{SUV} \mid \text{No}) = 0.6, \quad P(\text{Domestic} \mid \text{No}) = 0.6$$

**Posterior Probabilities:**

$$P(\text{Yes} \mid \text{Red, SUV, Domestic}) \propto 0.5 \cdot 0.6 \cdot 0.2 \cdot 0.4 = 0.024$$

$$P(\text{No} \mid \text{Red, SUV, Domestic}) \propto 0.5 \cdot 0.4 \cdot 0.6 \cdot 0.6 = 0.072$$

**Classification:** Since $P(\text{No}) > P(\text{Yes})$, the predicted class is:

$$\boxed{\text{No}}$$

# Problem 3: Naïve Bayes Classifier

## (a) Simplifying Assumption

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

$$P(X_1, X_2, \ldots, X_d \mid Y) = \prod_{i=1}^{d} P(X_i \mid Y)$$

This assumption reduces the computation required for estimating probabilities and simplifies the model.

## (b) Number of Parameters

Assume we have a binary classification problem with $d$ features and $k$ possible values for each feature.

**Without Simplifying Assumption:**

$$\text{Total Parameters} = 2 \cdot (k^d - 1)$$

**With Simplifying Assumption:**

$$\text{Total Parameters} = 2 \cdot d \cdot (k - 1)$$

For example, with $d = 3$ (features: Color, Type, Origin) and $k = 2$ (binary class labels: Yes/No):

$$\text{Without Assumption: } 2 \cdot (2^3 - 1) = 14$$

$$\text{With Assumption: } 2 \cdot 3 \cdot (2 - 1) = 6$$

## Why the Simplifying Assumption is Necessary

Without the simplifying assumption:

- The number of parameters grows exponentially with the number of features ($k^d$).
- High-dimensional data would require impractically large datasets to estimate probabilities reliably.
- Computationally, it becomes infeasible for large $d$ and $k$.

By assuming independence, the parameter count grows linearly with the number of features, making the model computationally efficient and suitable for practical applications.

# Question 4: Naïve Bayes Classification

**Problem:** We are tasked with classifying two short texts into one of three categories: **Physics**, **Biology**, or **Chemistry**, using the Naïve Bayes classifier. The probabilities of each category and key words are provided.

The two texts to classify are: - **Text A:** "The carbon atom is the foundation of life on earth." - **Text B:** "The carbon atom contains 12 protons."

We assume the following: 1. Words are conditionally independent given the category. 2. Words are stemmed to reduce them to their base form (e.g., "atoms" $\rightarrow$ "atom"). 3. Words not listed in the table are ignored.

—

## Step 1: Naïve Bayes Formula

For each category $c \in \{\text{Physics, Biology, Chemistry}\}$, the posterior probability is calculated as:

$$P(c \mid \text{text}) \propto P(c) \cdot \prod_{w \in \text{text}} P(w \mid c)$$

Where: - $P(c)$ is the prior probability of the category. - $P(w \mid c)$ is the likelihood of word $w$ given the category $c$.

The category with the highest posterior probability is the predicted classification.

—

## Step 2: Words to Use

After stemming, the relevant words from the texts are: - **Text A:** {carbon, atom, life, earth} - **Text B:** {carbon, atom, proton}

—

## Step 3: Classification of Text A

**Physics:**

$$P(\text{Physics} \mid \text{Text A}) \propto P(\text{Physics}) \cdot P(\text{carbon} \mid \text{Physics}) \cdot P(\text{atom} \mid \text{Physics}) \cdot P(\text{life} \mid \text{Physics}) \cdot P(\text{earth} \mid \text{Physics})$$

$$P(\text{Physics} \mid \text{Text A}) \propto 0.35 \cdot 0.005 \cdot 0.01 \cdot 0.001 \cdot 0.005 = 8.75 \times 10^{-10}$$

**Biology:**

$$P(\text{Biology} \mid \text{Text A}) \propto P(\text{Biology}) \cdot P(\text{carbon} \mid \text{Biology}) \cdot P(\text{atom} \mid \text{Biology}) \cdot P(\text{life} \mid \text{Biology}) \cdot P(\text{earth} \mid \text{Biology})$$

$$P(\text{Biology} \mid \text{Text A}) \propto 0.40 \cdot 0.03 \cdot 0.02 \cdot 0.1 \cdot 0.006 = 1.44 \times 10^{-6}$$

**Chemistry:**

$$P(\text{Chemistry} \mid \text{Text A}) \propto P(\text{Chemistry}) \cdot P(\text{carbon} \mid \text{Chemistry}) \cdot P(\text{atom} \mid \text{Chemistry}) \cdot P(\text{life} \mid \text{Chemistry}) \cdot P(\text{ea}$$

$$P(\text{Chemistry} \mid \text{Text A}) \propto 0.25 \cdot 0.03 \cdot 0.02 \cdot 0.008 \cdot 0.003 = 3.6 \times 10^{-8}$$

**Conclusion for Text A:** The highest posterior probability is for **Biology** ($1.44 \times 10^{-6}$), so:

$$\boxed{\text{Text A is classified as Biology}}$$

—

## Step 4: Classification of Text B

**Physics:**

$$P(\text{Physics} \mid \text{Text B}) \propto P(\text{Physics}) \cdot P(\text{carbon} \mid \text{Physics}) \cdot P(\text{atom} \mid \text{Physics}) \cdot P(\text{proton} \mid \text{Physics})$$

$$P(\text{Physics} \mid \text{Text B}) \propto 0.35 \cdot 0.005 \cdot 0.01 \cdot 0.05 = 8.75 \times 10^{-6}$$

**Biology:**

$$P(\text{Biology} \mid \text{Text B}) \propto P(\text{Biology}) \cdot P(\text{carbon} \mid \text{Biology}) \cdot P(\text{atom} \mid \text{Biology}) \cdot P(\text{proton} \mid \text{Biology})$$

$$P(\text{Biology} \mid \text{Text B}) \propto 0.40 \cdot 0.03 \cdot 0.02 \cdot 0.001 = 2.4 \times 10^{-7}$$

**Chemistry:**

$$P(\text{Chemistry} \mid \text{Text B}) \propto P(\text{Chemistry}) \cdot P(\text{carbon} \mid \text{Chemistry}) \cdot P(\text{atom} \mid \text{Chemistry}) \cdot P(\text{proton} \mid \text{Chemistry})$$

$$P(\text{Chemistry} \mid \text{Text B}) \propto 0.25 \cdot 0.03 \cdot 0.02 \cdot 0.05 = 7.5 \times 10^{-6}$$

**Conclusion for Text B:** The highest posterior probability is for **Physics** ($8.75 \times 10^{-6}$), so:

$$\boxed{\text{Text B is classified as Physics}}$$

—

## Final Answer(s):

1. **Text A: Biology**
2. **Text B: Physics**

# Question 5: Decision Trees for "Play Golf?"

## Dataset Overview

The dataset contains 14 examples with attributes: **Outlook**, **Temperature**, **Humidity**, **Windy**, and the target class **Play Golf?**.

—

## Tree 1: Using "Temperature" as the Root Node

**Entropy of Root Node:**

$$H(\text{Play Golf}) = -\left(\frac{9}{14} \cdot \log_2\left(\frac{9}{14}\right) + \frac{5}{14} \cdot \log_2\left(\frac{5}{14}\right)\right) \approx 0.940$$

**Weighted Average Entropy for Temperature:**

$$H(\text{Temperature}) \approx 0.911$$

**Information Gain for Temperature:**

$$IG(\text{Temperature}) = 0.940 - 0.911 = 0.029$$

**Tree Structure:**

$$\text{Temperature} \begin{cases} \text{Hot} & \text{[Split further]} \\ \text{Mild} & \text{[Split further]} \\ \text{Cool} & \text{[Split further]} \end{cases}$$

—

## Tree 2: Following the Decision Tree Learning Algorithm

**Step 1: Choose "Outlook" as the Root Node (Highest IG).**
**Tree Structure:**

$$\text{Outlook} \begin{cases} \text{Sunny} & \begin{cases} \text{Humidity} = \text{High} & \text{No} \\ \text{Humidity} = \text{Normal} & \text{Yes} \end{cases} \\ \text{Overcast} & \text{Yes} \\ \text{Rain} & \begin{cases} \text{Windy} = \text{True} & \text{No} \\ \text{Windy} = \text{False} & \text{Yes} \end{cases} \end{cases}$$

## Conclusion

Tree 1 uses "Temperature" as the root node and has low information gain. Tree 2 is more optimal as it follows the Decision Tree Learning algorithm and splits the data based on attributes with the highest information gain at each step.

1. [10 pts]

**1.3** ($\star\star$) Suppose that we have three coloured 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 that it came from the green box?

Representing the words as data

- 3 coloured boxes
  - → red, r
  - → blue, b
  - → green, g

⇒

| Box r | Box b | Box g |
|---|---|---|
| 3 apples | 1 apple | 3 apples |
| 4 oranges | 1 orange | 3 oranges |
| 3 limes | 0 limes | 4 limes |
| [Total 10 | 2 | 10] |

→ Table #Q1#

Q1 ⇓

| Box | P(Box) | Total # Fruit | P(Apple\|Box) | Weighted Probability $P(Box) \times P(Apple\|Box)$ |
|---|---|---|---|---|
| r | .2 | 10 | $\frac{3 \text{ apples}}{10 \text{ total fruits}} = .3$ | $.2 \times .3 = .06$ |
| b | .2 | 2 | $1/2 = .5$ | $.2 \times .05 = .10$ |
| g | .6 | 10 | $3/10 = .3$ | $.6 \times .3 = .18$ |

$\Sigma(\ ) = P(Apple)$

Question 1) If a box is chosen at random w/ probabilities

$p(r) = 0.2$

$p(b) = 0.2$ ($\Leftrightarrow$ Total $\Sigma = 1$)

$p(g) = 0.6$

AND a piece of fruit is removed from the box (w/ equal probability of selecting any of the items in the box)

→ Then what probability of selecting an apple?

Answer 1

$P(Apple) = 0.06 + 0.10 + 0.18$
$= 0.34$

Question 2) If we observe that the selected fruit is in fact an orange what is the probability that it came from the green box

↑ [orange]

Answer 2: Probability of Green Box Given Orange

| Table | Box | P(Box) | Oranges in Box | P(Orange\|Box) | Weighted Probability |
|---|---|---|---|---|---|
| | r | 0.2 | $4/10 = 0.4$ | 0.4 | $0.2 \times 0.4 = 0.08$ |
| | b | 0.2 | $1/2 = 0.5$ | 0.5 | $0.2 \times 0.5 = 0.10$ |
| | g | 0.6 | $3/10 = 0.3$ | 0.3 | $0.6 \times 0.3 = 0.18$ |

$\Rightarrow P(Orange) = 0.08 + 0.10 + 0.18$
$\equiv \underline{0.36}$

Now Applying Bayes' Theorem

$$P(g|Orange) = \frac{P(Orange|g) \cdot P(g)}{P(Orange)} = \frac{0.3 \times 0.6}{0.36} = \frac{.18}{.36} = \underline{0.5}$$

2. [20 pts] Given the following data set containing three attributes and one class, use naïve Bayes classifier to determine the class (Yes/No) of Stolen for a Red Domestic SUV.

| Example No. | Color | Type | Origin | Stolen? |
|---|---|---|---|---|
| 1 | Red | Sports | Domestic | Yes I |
| 2 | Red | Sports | Domestic | No I |
| 3 | Red | Sports | Domestic | Yes II |
| 4 | Yellow | Sports | Domestic | No II |
| 5 | Yellow | Sports | Imported | Yes III |
| 6 | Yellow | SUV | Imported | No III |
| 7 | Yellow | SUV | Imported | Yes IIII |
| 8 | Yellow | SUV | Domestic | No IV |
| 9 | Red | SUV | Imported | No √ |
| 10 | Red | Sports | Imported | Yes IIIII |

5 - Stolen
5 not s1

⇒ 5/10 stolen
5/10 not stolen

Restating attributes: Color → Red
Type → SUV
Origin → Domestic

Calculating likelihoods:
"Y/N"

| Class | Count (Red) | Total | P(Red \| Class) |
|---|---|---|---|
| Y | 3 | 5 | 3/5 = 0.6 |
| N | 2 | 5 | 2/5 = 0.4 |

likelihood for Color Red

| Class | Count (Red) | Total | P(Red \| Class) |
|---|---|---|---|
| Y | 1 | 5 | 1/5 = .2 |
| N | 3 | 5 | 3/5 = .6 |

likelihood for Type SUV

| Class | Count (Red) | Total | P(Red \| Class) |
|---|---|---|---|
| Y | 2 | 5 | 2/5 = .4 |
| N | 3 | 5 | 3/5 = 0.6 |

likelihood for Origin Domestic

Posterior Probabilities
[use Naive Bayes: Formula]

$P(Yes \mid Red, SUV, Domestic) \propto P(Yes) \cdot P(R \mid Yes) \cdot P(SUV \mid Yes) \cdot P(Domestic \mid Yes)$

$P(NO \mid \quad) \propto P(NO) \cdot P(R \mid N) \cdot P(SUV \mid NO) \cdot P(Domestic \mid No)$

⇒ for "Yes" :=

$\rightarrow P(Yes \mid Red, SUV, Domestic) \propto (0.5)(0.6)(0.2)(0.4) = 0.024$

<

⇒ for "NO" :=

$\rightarrow P(NO \mid Red, SUV, Domestic) \propto (0.5)(0.4)(0.6)(0.6) = 0.072$

Classify: Since $P(No \mid \ldots) > P(Yes \mid \ldots)$

⇒ prediction ≡ "NO"

## Question 3

3. [20 pts] This question is about naïve Bayes classifier. Please do the following:
   (a) State what is the simplifying assumption made by naïve Bayes classifier.
   (b) Given a binary-class classification problem in which the class labels are binary, the dimension of feature is *d*, and each attribute can take *k* different values. Please provide the numbers of parameters to be estimated with AND without the simplifying assumption. Briefly justify why the simplifying assumption is necessary.

*Full Writup on LateX

---

4. [25 pts] Assume we want to classify science texts into three categories—physics, biology and chemistry. The following probabilities have been estimated from analyzing a corpus of pre-classified web-pages gathered from Yahoo.

**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.***

- For each category $c \in \{\text{Pysics}, \text{Bio}, \text{chemistry}\}$ we need posterior probability:

  $P(c|\text{text}) \propto P(c) \cdot \prod_{w \in \text{text}} P(w|c)$

  w/ $P(c)$ prior prob. of category
  $P(w|c)$ liklihood of word w given the category c

Ans: A Biology
B Physics

Steps
- Classify text into Categories
  - Physics
  - Biology  w/ Naive Bayes
  - Chemistry  Identifier

Text A
→ Stemming: {Carbon, atom, life, earth}

Text B
→ {carbon atom proton}

Classify Category

Physics $\propto_{\dots} = .35 \cdot .005 \cdot .01 \cdot .001 \cdot .005$
$\approx 8.75 \times 10^{-10}$

Biology $\propto_{\dots} = .40 \cdot .03 \cdot .02 \cdot .1 \cdot .006$
$\approx 1.44 \times 10^{-6}$ ← Largest

Chem $\propto \dots = .25 \cdot .03 \cdot .02 \cdot .008 \cdot .003$
$\approx 3.6 \times 10^{-8}$

sim for B (see LateX !!)

5. Consider dataset:

| 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 |

From the classified examples in the above table, construct two decision trees (by hand) for the classification "Play Golf." For the first tree, use Temperature as the root node. (This is a really bad choice.) Continue the construction of tree as discussed in class for the subsequent nodes using information gain. Remember that different attributes can be used in different branches on a given level of the tree. For the second tree, follow the Decision Tree Learning algorithm described in class. At each step, choose the attribute with the highest information gain. Work out the computations of information gain by hand and draw the decision tree.

Answer

- need 2 D Trees for "Play Golf"
  - A. Temperatur = root
  - B. DT Learning Algo
    - ⇒ highest info gain 1 step!

1) analyze data - 14 examples

- outlook - sun / overcast / rain
- temperature - hold / mild / cool
- humidity - high / normal
- windy - true / false
- play golf - yes / no

Play Golf
Y - 9
N - 5

2) Calculate Entropy

$$H(S) = -\sum_{i=1}^{k} P_i \cdot \log_2(P_i)$$

$P(Yes) = 9/14$

$P(No) = 5/14$

$H(\text{Play Golf}) = -\left(\frac{9}{14} \cdot \log_2\left(\frac{9}{14}\right) + \frac{5}{14}\log_2\left(\frac{5}{14}\right)\right)$

$\approx 0.94028$

Step 3) Obtaining tree 1 organization

- Temperature is the root node!

→ Split via Temp:

- Hot $|S| = 4$ Yes $= 2$ No $= 2$
- Mild $|S| = 6$ Yes $= 4$ No $= 2$
- Cool $|S| = 4$ Yess $= 3$ No $= 1$

→ entropy of each subset

$H(\text{Hot}) = -\left(\frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right) + \frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right)\right)$

$\approx 1.0$ ($\equiv 1$ [scientific Notation permitted])

$H(\text{Mild}) = -\left(\frac{4}{6} \cdot \log_2(4/6) + 2/6 \cdot \log_2(2/6)\right)$

$\approx 0.91829...$

$H(\text{Cool}) = -\left(3/4 \cdot \log_2\left(\frac{3}{4}\right) + 1/4 \log_2(1/4)\right)$

$\approx 0.81127...$

⇒ Weighte Avg Entropy (info gain for temp)

$H(\text{Temp}) = \frac{4}{14} \cdot (1.0) + \frac{6}{14}(0.918) + \frac{4}{14}(.811)$

$\approx 0.029$

Tree A

- Temp
  - Hot
  - Mild
  - Cool

Tree B

- Outlook
  - Sunny
    - Humidity
      - High → No
      - Low → Yes
  - Overcast → Yes
  - Rain
    - Windy
      - True → No
      - False Yes

"Play Golf" ANSWERS

Step 4 Tree B [Breakdown]

I.G for Attrib.

| | |
|---|---|
| Outlook | .246 |
| Temp | .029 |
| Humidity | .151 |
| Winty | .048 |

"Largest" ≡ "Best" ⇒ Root: Outlook

[Split by]

| Outlook: | $\lvert S\rvert = 5$ | $\lvert S\rvert = 4$ | $\lvert S\rvert = 5$ |
|---|---|---|---|
| Yes | 2 | 4 | 3 |
| No | 3 | 0 | 2 |

subtree: Humidity → High ⇒ No Normal ⇒ Yes

subtree: Overcast → Yes [only]

subtree: rain → windy = True ⇒ No

windy = False ⇒ Yes

**CSCE 421: Machine Learning (Spring 2025)**

**Homework #1**

**Due 2/4/2025 11:59 pm**

---

- You need to submit a report in PDF to Canvas.
- Please name your PDF report "HW#_FirstName_LastName.pdf". Please submit your PDF file directly to Canvas (i.e., do not include the PDF file into a ZIP file).
- All students are highly encouraged to typeset their reports using Word or LaTeX. In case you decide to hand-write, please make sure your answers are clearly readable in scanned PDF.
- Unlimited number of submissions are allowed and the latest one will be timed and graded.
- Please read and follow submission instructions. No exception will be made to accommodate incorrectly submitted files/reports.
- Please start your submission to Canvas at least 15-30 minutes before the deadline, as there might be latency. We do NOT accept E-mail submissions.

---

1. [10 pts]

**1.3** ($\star\star$) Suppose that we have three coloured 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 that it came from the green box?

2. [20 pts] Given the following data set containing three attributes and one class, use naïve Bayes classifier to determine the class (Yes/No) of Stolen for a Red Domestic SUV.

| 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 |

3. [20 pts] This question is about naïve Bayes classifier. Please do the following:
   (a) State what is the simplifying assumption made by naïve Bayes classifier.
   (b) Given a binary-class classification problem in which the class labels are binary, the dimension of feature is *d*, and each attribute can take *k* different values. Please provide the numbers of parameters to be estimated with AND without the simplifying assumption. Briefly justify why the simplifying assumption is necessary.

4. [25 pts] Assume we want to classify science texts into three categories—physics, biology and chemistry. The following probabilities have been estimated from analyzing a corpus of pre-classified web-pages gathered from Yahoo.

**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.***

5. [25 pts] Consider the following table of observations:

| 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 |

From the classified examples in the above table, construct two decision trees (by hand) for the classification "Play Golf." For the first tree, use Temperature as the root node. (This is a really bad choice.) Continue the construction of tree as discussed in class for the subsequent nodes using information gain. Remember that different attributes can be used in different branches on a given level of the tree. For the second tree, follow the Decision Tree Learning algorithm described in class. At each step, choose the attribute with the highest information gain. Work out the computations of information gain by hand and draw the decision tree.