# CMSC5724: Quiz 1

**Name:** **Student ID:**

**Problem 1 (50%).** Consider the set of training data shown below. Here, $A, B, C$ are attributes, and $D$ is the class label.

| $A$ | $B$ | $C$ | $Y$ |
|---|---|---|---|
| 1 | 1 | 1 | y |
| 1 | 0 | 1 | y |
| 0 | 0 | 1 | y |
| 1 | 1 | 0 | y |
| 0 | 1 | 1 | n |
| 1 | 1 | 1 | n |
| 0 | 0 | 0 | n |
| 0 | 1 | 0 | n |

Suppose that we consider only decision trees in the form described in Figure 1: there are 3 nodes (i.e., a root and two leaves) where $X$ is an attribute ($A$, $B$, or $C$) and $v$ is an integer chosen from $\{0, 1\}$. Give a decision tree conforming to the template whose empirical error is the smallest (you do not need to explain how the tree is found).

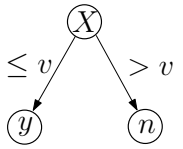


Figure 1

**Solution.** $X = B$ and $v = 0$.

**Problem 2 (50%).** Consider the same training data again, repeated here for your convenience.

| $A$ | $B$ | $C$ | $Y$ |
|---|---|---|---|
| 1 | 1 | 1 | y |
| 1 | 0 | 1 | y |
| 0 | 0 | 1 | y |
| 1 | 1 | 0 | y |
| 0 | 1 | 1 | n |
| 1 | 1 | 1 | n |
| 0 | 0 | 0 | n |
| 0 | 1 | 0 | n |

Apply naive Bayes classification to predict the label of an unseen record with $A = 1$, $B = 0$, $C = 0$. You must show the details of your reasoning.

**Solution.** We need to compare the values of $\mathbf{Pr}[(1, 0, 0) \mid Y = y] \cdot \mathbf{Pr}[Y = y]$ and $\mathbf{Pr}[(1, 0, 0) \mid Y = n] \cdot \mathbf{Pr}[Y = n]$. The estimation for both $\mathbf{Pr}[Y = y]$ and $\mathbf{Pr}[Y = n]$ is $1/2$. So it remains to compare $\mathbf{Pr}[(1, 0, 0) \mid Y = y]$ and $\mathbf{Pr}[(1, 0, 0) \mid Y = n]$. Following the assumption of naive Bayes classification,

we estimate the two terms as follows:

$$\mathbf{Pr}[(1,0,0) \mid Y = y] = \mathbf{Pr}[A = 1 \mid Y = y] \cdot \mathbf{Pr}[B = 0 \mid Y = y] \cdot \mathbf{Pr}[C = 0 \mid Y = y] = \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{4} = \frac{3}{32}$$

$$\mathbf{Pr}[(1,0,0) \mid Y = n] = \mathbf{Pr}[A = 1 \mid Y = n] \cdot \mathbf{Pr}[B = 0 \mid Y = n] \cdot \mathbf{Pr}[C = 0 \mid Y = n] = \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{32}$$

Based on the above estimation, we predict the label to be $y$.