CMSC5724: Quiz 1

Name: **Student ID:**

Problem 1 (30%). Consider the training data shown below. Here, A and B are attributes, and Y is the class label.

A	B	Y
2	3	y
6	1	y
1	12	y
3	9	y
11	15	n
7	13	n
4	8	n
9	10	n

Suppose that we consider only decision trees in the form described in Figure 1: there are 3 nodes (i.e., a root node and two leaves) where X is an attribute (either A or B) and v is an integer chosen from $\{0, 1, ..., 15\}$. Give a decision tree conforming to the template whose empirical error is at most 0.125.

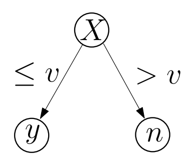
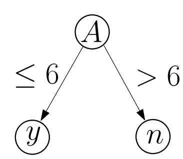


Figure 1

Answer. One possible decision tree is shown below.



Problem 2 (30%). Use the generalization theorem (in Lecture Notes 1) to prove that the generalization error of your decision tree in Problem 1 is at most 0.73. Again, we consider only the decision trees conforming to the template in Figure 1. Your estimate should be correct with probability at least 90%.

Answer. Les S be the training set given in Problem 1 and $\mathcal{H}$ be the set of classifiers that can possibly be returned. Denote by h the decision tree we found in Problem 1. As X has two choices (A and B) and v has 16 choices, we know $|\mathcal{H}| = 32$. Our decision tree in Problem 1 has empirical error $err_S(h) = 0.125$.

According to the generalization theorem, with probability at least $1-\delta$, we have

$$\begin{aligned} err_{\mathcal{D}}(h) &\leq err_S(h) + \sqrt{\frac{\ln(1/\delta) + \ln|\mathcal{H}|}{2|S|}} \\ &= 0.125 + \sqrt{\frac{\ln(1/\delta) + \ln 32}{16}}. \end{aligned}$$

By setting $\delta = 0.1$, we know with probability at least 0.9,

$$err_{\mathcal{D}}(h) \leq 0.125 + \sqrt{\frac{\ln(1/0.1) + \ln 32}{16}} \leq 0.73.$$

Problem 3 (40%). Consider following training data, where A, B, C are attributes, and Y is the class label.

A	B	C	Y
1	1	1	y
1	0	1	y
0	1	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 $A = 1$, $B = 1$, $C = 0$. You must show the details of your reasoning.

Answer. By Bayes Theorem

$$\mathbf{Pr}[Y = y \mid A = 1, B = 1, C = 0] = \frac{\mathbf{Pr}[A = 1, B = 1, C = 0 \mid Y = y] \cdot \mathbf{Pr}[Y = y]}{\mathbf{Pr}[A = 1, B = 1, C = 0]}$$

and

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

To know which fraction is bigger, it is sufficient to estimate their numerators:

$$\begin{aligned} & \mathbf{Pr}[A = 1, B = 1, C = 0 \mid Y = y] \cdot \mathbf{Pr}[Y = y] \\ = \; & \mathbf{Pr}[A = 1 \mid Y = y] \cdot \mathbf{Pr}[B = 1 \mid Y = y] \cdot \mathbf{Pr}[C = 0 \mid Y = y] \cdot \mathbf{Pr}[Y = y] \\ \text{(estimate)} \;\; = \; & \frac{3}{4} \times \frac{3}{4} \times \frac{1}{4} \times \frac{1}{2} \\ = \; & \frac{9}{128}. \end{aligned}$$

$$\begin{aligned} & \mathbf{Pr}[A = 1, B = 1, C = 0 \mid Y = n] \cdot \mathbf{Pr}[Y = n] \\ = \; & \mathbf{Pr}[A = 1 \mid Y = n] \cdot \mathbf{Pr}[B = 1 \mid Y = n] \cdot \mathbf{Pr}[C = 0 \mid Y = n] \cdot \mathbf{Pr}[Y = n] \\ \text{(estimate)} \;\; = \; & \frac{1}{4} \times \frac{3}{4} \times \frac{2}{4} \times \frac{1}{2} \\ = \; & \frac{6}{128}. \end{aligned}$$

We thus conclude that $\mathbf{Pr}[Y = y \mid A = 1, B = 1, C = 0] > \mathbf{Pr}[Y = n \mid A = 1, B = 1, C = 0]$. The predicted label is y.