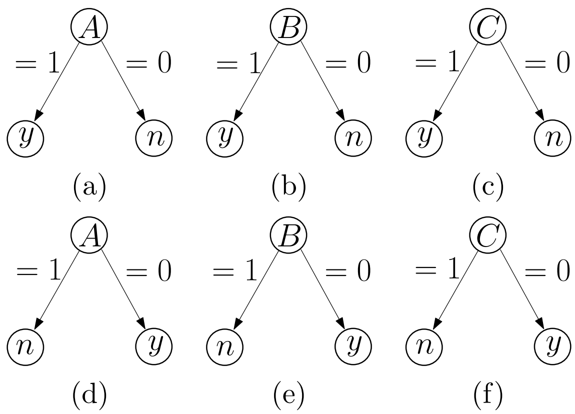
CMSC5724: Quiz 1

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

A	B	C	Y
1	1	1	y
0	1	1	y
0	0	1	y
1	1	0	y
1	0	1	n
1	1	1	n
0	0	0	n
1	0	0	n

Suppose that we consider only decision trees each having 3 nodes — namely, a root node and two leaves — where one leaf has label 'y' and the other has label 'n'. Give the decision tree with the best empirical error. You need to explain your reasoning.

Answer. For the given input, there are only 6 possible decision trees having 3 nodes, which are:



Among them, the decision tree (b) has the lowest empirical error 1/4 and, hence, is the answer.

Problem 2 (40%). Use the generalization theorem (in Lecture Notes 1) to estimate the generalization error of your decision tree in Problem 1. Again, we consider only the decision trees with 3 nodes where one leaf has label 'y' and the other has label 'n'. Your estimate should be correct with probability at least 99%.

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 best decision tree we found in Problem 1. From the above solution, we know $|\mathcal{H}| = 6$ and the empirical error $err_S(h) = 1/4$.

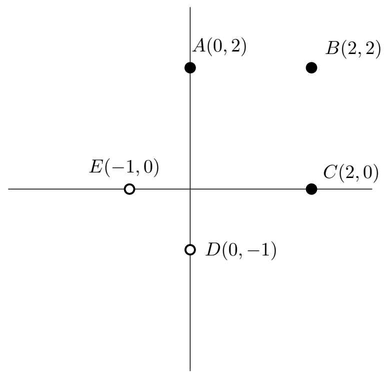
According to the generalization theorem, with probability at least $1 - \delta$, we have

$$\begin{aligned} err_{\mathcal{D}}(h) &\leq err_S(h) + \sqrt{\frac{\ln(1/\delta) + \ln|\mathcal{H}|}{2|S|}} \\ &\leq 1/4 + \sqrt{\frac{\ln(1/\delta) + \ln 6}{16}}. \end{aligned}$$

By setting $\delta = 0.01$, we know with probability at least 0.99,

$$err_{\mathcal{D}}(h) \leq 1/4 + \sqrt{\frac{\ln(1/0.01) + \ln 6}{16}}.$$

Problem 3 (30%). The following figure shows a set of 5 points. Use the Perceptron algorithm to find a line that (i) crosses the origin and (ii) separates the black points from the white ones. Recall that Perceptron starts with a vector $\boldsymbol{w} = \mathbf{0}$ and iteratively adjusts it using a violation point. You need to show the value of $\boldsymbol{w}$ after every adjustment.



Answer: Without loss of generality, assume that the black points have label 1 while the white ones have label -1. At the beginning, $\boldsymbol{w} = (0, 0)$. We use $\boldsymbol{A}$ to denote the vector form of A. Define $\boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ and $\boldsymbol{E}$ similarly.

Iteration 1. Since $\boldsymbol{A}$ does not satisfy $\boldsymbol{w} \cdot \boldsymbol{A} > 0$, update $\boldsymbol{w}$ to $\boldsymbol{w} + \boldsymbol{A} = (0, 0) + (0, 2) = (0, 2)$.

Iteration 2. Since $\boldsymbol{C}$ does not satisfy $\boldsymbol{w} \cdot \boldsymbol{A} > 0$, update $\boldsymbol{w}$ to $\boldsymbol{w} + \boldsymbol{C} = (0, 2) + (2, 0) = (2, 2)$.

Iteration 3. No more violation points. So we have found a separation line $2x + 2y = 0$.