MSML603: Midterm Solution (PCS05) **Fall 2024**
Name:

Please sign the following honor pledge. Ten (10) points will be taken off if not signed.

"I pledge on my honor that I have not given or received any unauthorized assistance on this exam."

Signature:

PLEASE EXPLAIN YOUR ANSWERS THOROUGHLY. YOU WILL BE GRADED BASED ON THE EXPLANATIONS AND STEPS YOU TAKE, NOT JUST ON THE FINAL ANSWERS.

1. Suppose that radiation measurements are taken inside a reactor to detect a possible issue at a nuclear plant. The logarithm of the radiation measurement is denoted by X. When there is an issue with the reactor, $X \sim \mathcal{N}(5, 5^2)$. On the other hand, when the reactor is working properly without any issue $X \sim \mathcal{N}(-10, 5^2)$. Recall that, for a Gaussian random variable $Z \sim \mathcal{N}(\mu, \sigma^2)$, its probability density function is given by

$$f_Z(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right), \quad z \in \mathbb{R} .$$

On a given day, before a measurement is taken, the probability that there is an issue at the nuclear plant is 0.05. Based on the measurement X, we need to choose either `'possible issue at the reactor'` (I) or `'reactor working properly'` (NI).

(a) Design a Bayesian decision rule that minimizes the probability of making a mistake. What action would the decision rule choose when the measurements is $X = 2$? (13 pts)

Ans: In order to minimize the probability of making a mistake, we should choose the hypothesis that has larger posterior probability.

$$\mathbf{P}(I \mid X = x) \propto \mathbf{P}(I)\, p_I(x) = 0.05 \frac{1}{\sqrt{2\pi}5} \exp\left(-\frac{(x-5)^2}{50}\right)$$
$$\mathbf{P}(NI \mid X = x) \propto \mathbf{P}(NI)\, p_{NI}(x) = 0.95 \frac{1}{\sqrt{2\pi}5} \exp\left(-\frac{(x-(-10))^2}{50}\right)$$

We should choose 'NI' if $\frac{\mathbf{P}(NI \mid X=x)}{\mathbf{P}(I \mid X=x)} > 1$ or, equivalently,

$$19 \exp\left(\frac{-(x+10)^2 + (x-5)^2}{50}\right) = 19 \exp\left(\frac{-6x-15}{10}\right) > 1 .$$

Taking the natural logarithm, we choose 'NI' if

$$\log(19) - \frac{6x+15}{10} > 0 \text{ or, equivalently } x < \frac{5}{3}\log(19) - \frac{5}{2} = 2.41 .$$

(b) The cost of mistakenly declaring that there is an issue with the reactor when there is no problem is 5. On the other hand, the cost of missing an issue with the reactor is 20. There is no cost when a correct decision is made. Design a Bayesian decision rule that minimizes the overall risk. (13 pts)

Ans: The conditional risks are given by

$$R(NI \mid X = x) = 20\mathbf{P}(I \mid X = x) \text{ and } R(I \mid X = 2) = 5\mathbf{P}(NI \mid X = x) .$$

Since we should 'NI' if $\frac{R(I \mid X=x)}{R(NI \mid X=2)} = \frac{5\mathbf{P}(NI \mid X=x)}{20\mathbf{P}(I \mid X=x)} > 1$, which is equivalent to $x < \frac{5}{3}\log(19/4) - \frac{5}{2} = 0.097$.

2. We want to approximate the original samples with three features using principal component analysis (PCA). Assume that the scatter matrix obtained from the samples is

$$S = \begin{bmatrix} 7.5 & 2.5 & 0 \\ 2.5 & 7.5 & 0 \\ 0 & 0 & 2 \end{bmatrix} .$$

(a) Determine the first principal component $\mathbf{w}$. (13 pts)

Ans: In order to find the first principal component, we first need to find the eigenvalues of the scatter matrix using the characteristic equation.

$$\begin{aligned} \det(\lambda\mathbf{I} - S) &= ((\lambda - 7.5)^2 - 2.5^2)(\lambda - 2) \\ &= (\lambda^2 - 15\lambda + 50)(\lambda - 2) = (\lambda - 10)(\lambda - 5)(\lambda - 2) = 0 \end{aligned}$$

Hence, the three eigenvalues are $\lambda_1 = 10$, $\lambda_2 = 5$ and $\lambda_3 = 2$. The first principal component is an eigenvector $\mathbf{v}^1$ associated with λ_1, which satisfies $S\mathbf{v}^1 = 10\mathbf{v}^1$.

$$\begin{aligned} 7.5v_1^1 + 2.5v_2^1 = 10v_1^1 &\Rightarrow v_1^1 = v_2^1 \\ 2v_3^1 = 10v_3^1 &\Rightarrow v_3^1 = 0 \end{aligned}$$

From the above equations and normalizing the eigenvector, we obtain $\mathbf{v}^1 = [1/\sqrt{2}\ 1/\sqrt{2}\ 0]^T = \mathbf{w}$.

(b) Find the approximation of a sample $\mathbf{x} = [1.2\ 1.7\ \text{-}1.3]^T$ using the first principal component from part (a), i.e., a vector $\tilde{\mathbf{x}}$ in the line $\{\alpha\mathbf{w} \mid \alpha \in \mathbb{R}\}$ which is closest to $\mathbf{x}$. (12 pts)

Ans: The projection of $\mathbf{x}$ onto the linear span of $\mathbf{w}$ is given by

$$\tilde{\mathbf{x}} = \frac{\langle \mathbf{x}, \mathbf{w} \rangle}{\|\mathbf{w}\|^2}\mathbf{w} = [1.45\ \ 1.45\ \ 0]^T .$$

3. We wish to find a margin perceptron for binary classification using a linearly separable dataset with labels +1 and -1. Suppose that the margin perceptron is given by the hyperplane $\{\mathbf{x} \in \mathbb{R}^3 \mid [1\ \text{-}0.5\ 2]\mathbf{x} = 2\}$. Assume that the hyperplane correctly separates all samples.

(a) Suppose that one of the samples in the dataset used for training is $\mathbf{x} = [1\ \text{-}3\ 2]^T$. Compute the distance from the hyperplane to the sample $\mathbf{x}$ and determine its label. (13 pts)

Ans: Let $\mathbf{w} = [1\ \text{-}0.5\ 2]^T$, and define $g : \mathbb{R}^3 \to \mathbb{R}$, where $g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} - 2$. Then, the hyperplane is given by $\{\mathbf{x} \in \mathbb{R}^3 \mid g(\mathbf{x}) = 0\}$. Then, from problem 1 in homework assignment #4, the distance between the given sample $[1\ \text{-}3\ 2]^T$ and the hyperplane is given by $|g([1\ \text{-}3\ 2]^T)|/\|\mathbf{w}\| = 4.5/2.29 = 1.96$.

(b) Compute the margin of the provided margin perceptron. (12 pts)

Ans: From the discussion on support vector machine (Lecture 6), the margin of a margin perception given by $(\mathbf{w}, b)$ is equal to $2/\|\mathbf{w}\| = 0.873$.

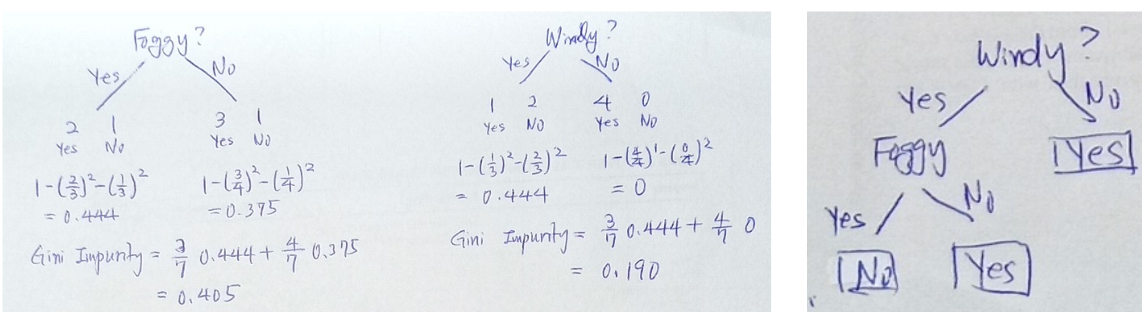
4. We are interested in predicting whether or not a flight will depart on time. In order to help us better predict the outcome, we want to build a decision tree using the answers to the questions shown in Table 4. Here, we will use the first two questions `'foggy?'` and `'windy?'` to build the decision tree.

foggy?	windy?	depart on time?
Yes	No	Yes
No	No	Yes
No	Yes	No
Yes	No	Yes
Yes	Yes	No
No	No	Yes
No	Yes	Yes

Table 1: Dataset for problem 4.

(a) Use Gini impurity and build a decision tree. (12 pts)

Ans: The Gini impurity for `'foggy?'` is $\frac{3}{7} \times 0.444 + \frac{4}{7} \times 0.375 = 0.405$. Similarly, that for `'windy?'` is $\frac{3}{7} \times 0.4444 + \frac{4}{7} \times 0 = 0.190$. Hence, the decision tree should ask the second question `'windy?'` at the root node. However, the second question `'foggy?'` only needs to be asked to the group 'yes' to the first question since when it was not windy, the flight always departed on time. This is shown in Figure 1(a). The constructed decision tree is shown in Figure 1(b).



(a) (b)

Figure 1: (a) Gini impurity, (b) decision tree.

(b) Use the information gain and build a decision tree. (12 pts)

Ans: Following similar steps in part (a), the information gain for `'foggy?'` is 0.006, and that for `'windy?'` is 0.470. Thus, the second question should be asked at the root, which results in the same decision tree. This is shown in Figure 2.

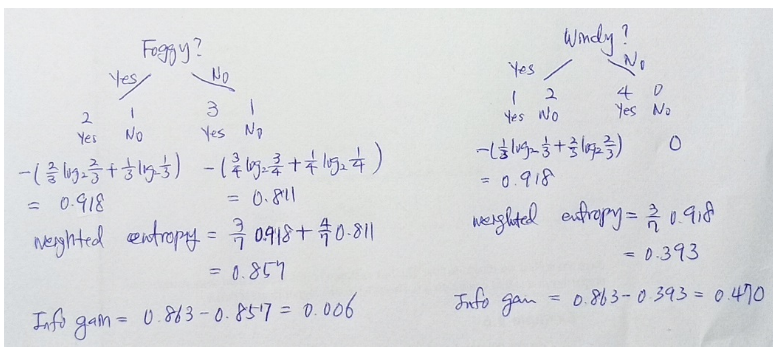


Figure 2: Information gain for problem 4(b).