**CSE475 Midterm 1, Due: 11:59 pm, Monday 03/03/2025**

**Please note that you are encouraged to typeset your solutions using either LATEX or Microsoft Word, and produce a PDF file for submission. Alternatively, you can scan your handwritten solutions and produce a PDF file. Unreadable/illegible solutions will not be graded. You need to submit an electronic version (in PDF form) on the canvas. You should name your file using the format CSE475-Midterm1-LastName-FirstName.**

1. [10pts] True-or-False questions. Please answer either "T" (True) or "F" (False) for the following statements.

[ ] The probability density function of a continuous random variable is denoted by $f(x)$, and we know $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$. It is possible that $f(x) > 1$ for some $x$.

[ ] The decision boundary of K Nearest Neighbors classifier is always linear (in other words, the decision boundary is always straight lines).

[ ] Different distances measures (such as Euclidean distance, L1 distance etc.) assign different decision boundaries to K-nearest neighbor classifier.

[ ] Classification accuracy (that is, the fraction of test points correctly classified by a classification algorithm) can be used to evaluate a Linear Regression model.

2. [20pts] We will train a Naïve Bayes classifier for a binary classification problem with four categorical features, $X_1, X_2, X_3, X_4$, and each feature can take values in $\{1,2,3\}$. We use $Y$ to denote the class label, and $Y$ is either 1 or -1. Suppose we have enough training data to estimate every feature likelihood and the class prior probability.

1) [5pts] What is the minimum number of parameters of the Naïve Bayes classifier which should be estimated before we can use the classifier to predict the class label of any given features?

2) [15pts] Please list the parameters in the set of parameters of minimum size (such minimum size is the answer to part 1) ).

Hint: The parameters of a Naïve Bayes classifier are feature likelihoods and the class prior probability which are estimated from training data. Please note that we only need to identify the set of parameters of minimum size containing only the necessary parameters. For example, if $P(Y = 1)$ is in the set, then $P(Y = 0)$ is not in the set because it can be computed by $1 - P(Y = 1)$. For another example, if $P(X_1 = 1|Y = 1)$ and $(X_1 = 2|Y = 1)$ are in this set, then $P(X_1 = 3|Y = 1)$ is not in the set because it can be computed by $1 - P(X_1 = 1|Y = 1) -$

$P(X_1 = 2|Y = 1)$. Please also note that such minimum set is not unique, and you only need to work on a particular one.

3. [15pts] A student is late for class. It could be because the bus arrived late, or the student overslept. The reason for oversleeping could be the alarm did not ring. There are four random variables involved in this problem, which are $A$ (Alarm ring), $O$ (Overslept), $B$ (Bus Late), $L$ (Late for Class). Each of them can take value of either 1 (corresponding to "Yes") or 0 (corresponding to "No").

[1] Construct a Bayesian Network (BN) for the above problem with the four random variables as nodes (please plot the BN).

[2] Please factorize the joint probability $P(A, O, B, L)$ using the structure of the BN constructed in part [1].

[3] We want to estimate the joint probability $P(A, O, B, L)$ for all possible values of $A, O, B, L$. Please list all the parameters that are needed. Hint: Parameters of a BN are prior probabilities and conditional probabilities. Please list a minimum set of necessary parameters which do not include parameter(s) that can be deduced from other parameters.

[4] Write an expression using the parameters from part [3] that represents the probability of reaching the class on time if the student overslept.

4. [10pts] Consider a one-dimensional two-class classification problem, with the class-conditional PDFs for the two classes ($\omega_1$ and $\omega_2$) being the normal densities $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively. Note that these two PDFs have the same variance $\sigma^2$. $N(\mu, \sigma^2)$ denotes the normal density defined by

$$p(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}.$$

The losses are defined in the same way as in the lecture, and the definition is recapped below for your convenience: $\lambda_{ij}$ is the loss incurred when classifying a sample as $\omega_i$ while the true class label is $\omega_j$. We let $\mu_1 = -1, \mu_2 = 1, \sigma^2 = 0.5$ for the following questions.

[1] If we do minimum-error-classification with simple 0-1 losses (i.e., $\lambda_{11} = \lambda_{22} = 0, \lambda_{21} = \lambda_{12} = 1$), the minimum-error-rate classification (by Bayes decision rule) will decide class label $\omega_1$ for feature $x$ such that $x < a$, and this is the Bayesian decision boundary. What is the value of $a$ if the classes have equal prior probabilities ($P(\omega_1) = P(\omega_2) = \frac{1}{2}$)?

[2] If the classes have equal prior probabilities ($P(\omega_1) = P(\omega_2) = \frac{1}{2}$), but the losses are given by $\lambda_{11} = \lambda_{22} = 0, \lambda_{12} = 1, \lambda_{21} = 2$. The minimum-error-rate classification (by Bayes decision rule) still decides class label $\omega_1$ for feature $x$ such that $x < a$. What is the value of $a$?

[3] If we use the same losses as part [2]: $\lambda_{11} = \lambda_{22} = 0$, $\lambda_{12} = 1$, $\lambda_{21} = 2$, but change the prior to $P(\omega_1) = \frac{1}{3}$. The minimum-error-rate classification (by Bayes decision rule) still decides class label $\omega_1$ for feature $x$ such that $x < a$. What is the value of $a$?

5. [10pts]. Consider a one-dimensional two-class classification problem with class-conditional PDFs and class prior probabilities as follows:

$$p(x|\omega_1) = \begin{cases} 4x^3 & x \in [0,1] \\ 0 & \text{otherwise} \end{cases}, \quad p(x|\omega_2) = \begin{cases} 6x - 6x^2 & x \in [0,1] \\ 0 & \text{otherwise} \end{cases},$$

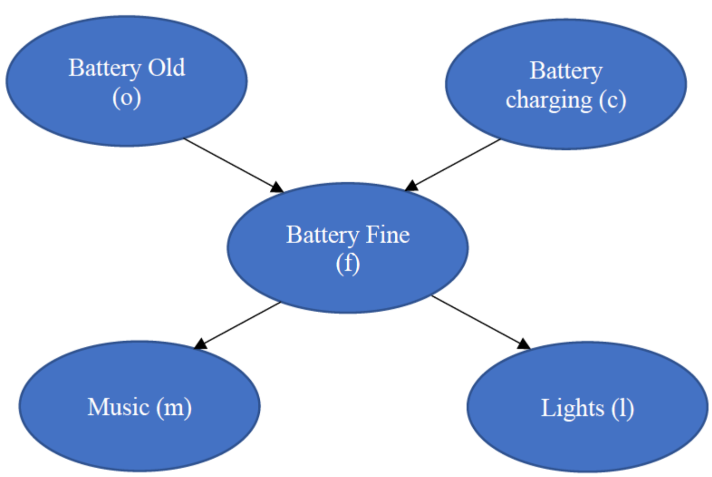
$P(\omega_1) = \frac{9}{10}$ (and thus $P(\omega_2) = \frac{1}{10}$). $\omega_1, \omega_2$ are two class labels.

[1] Specify the condition on $x$ (in terms of $x < a$ or $x > a$ for a specific $a$) such that the minimum-error-rate classification (by Bayes decision rule) will decide class label $\omega_1$ for feature $x$.

[2] What is the average error using the Bayes decision rule? Note that the average error should be computed by $P(error) = \mathrm{E}[P(error \mid x)] = \int P(error \mid x) p(x) \mathrm{d}x$.

6. [10pts] Given three random variables $A, B, C$, suppose that $P(A|B) = P(A|C)$. Is it always true that $P(B|A) = P(C|A)$? If it is always true, please derive it. Otherwise, please give a counter-example.

7. [15pts] Consider the following simple Bayesian Network (BN), where all the nodes are assumed to be binary random variables and each of them can take value of either 1 or 0.



Suppose we use the following abbreviations: $o = \text{Battery Old}, c = \text{Battery charging}, f = \text{Battery Fine}, m = \text{Music}, l = \text{Lights}$.

Find an expression for the probability that lights are on given that batter is old, i.e. $P(l = 1|o = 1)$, using only the following parameters (probabilities):

$$P(o = 1); P(c = 1); P(f = 1|o = 0, c = 0); P(f = 1|o = 0, c = 1); P(f = 1|o = 1, c = 0);$$
$$P(f = 1|o = 1, c = 1); P(l = 1|f = 0); P(l = 1|f = 1); P(m = 1|f = 0); P(m = 1|f = 1).$$

Please show how this expression is derived. (Hint: it is fine if some given parameters are not used. You can use the fact that $o, c$ are independent.)

8. [10pts] Consider the following data of 100 patients who are classified as healthy or sick, where "Count" indicates the number of patients with corresponding features.

| Sick | | |
|---|---|---|
| Fever | Headache | Count |
| T | T | 14 |
| T | F | 6 |
| F | T | 7 |
| F | F | 3 |

| Healthy | | |
|---|---|---|
| Fever | Headache | Count |
| T | T | 2 |
| T | F | 3 |
| F | T | 7 |
| F | F | 58 |

Using the above data and the Naïve Bayes algorithm, calculate the probability of a patient being sick if
[1] $\text{Headache} = \text{T}$ and $\text{Fever} = \text{F}$
[2] $\text{Headache} = \text{F}$ and $\text{Fever} = \text{T}$.

Hint: the Naïve Bayes algorithm assumes that the Naïve Bayes Assumption holds, i.e. the two features, Fever and Headache, are conditionally independent given Sick or Healthy.