March 22, 2024
IIT Jodhpur

Minor-2

CSL2050 - Pattern Recognition and Machine Learning

NOTE:

This is Question-cum-Answer sheet. Maximum Points: 50, Total Questions: 6, Total Pages: 3 (6 sides), Total Time: 1 Hour. **If there is anything not clear in the problems, go ahead with your own assumptions but state them clearly. No doubts will be entertained during the exam.** Be precise and write the answer in the box provided. Verbosity will be penalized. Use the other answer sheet for rough work and submit both.

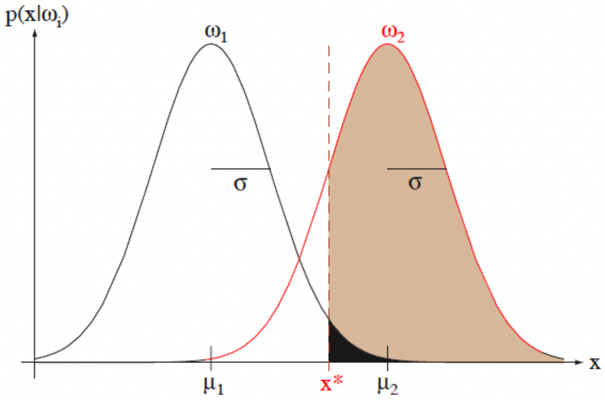
Name: **Roll Number:** **Signature:**

1. Let X be a discrete random variable with support $\{-1, 0, 1\}$. Suppose P_{-1}, P_0 and P_1 denote probability at -1, 0 and 1, respectively. Then, compute the expected value and variance of this PMF in terms of P_{-1}, P_0 and P_1.
   (a) **(2 pts)** E(X) (Only final answer)

   (b) **(3 pts)** Var(X) (Only final answer)

   (c) **(5 pts)** Find a PMF for X for which Var(X) has the maximum possible variance. What value of variance do you get in this case?

2. Suppose the probability of a test sample x is being classified as class ω_i is given by $P(\omega_i|x)$. Now, consider a classifier $x = x^*$ shown next:



Assuming ω_2 as positive and ω_1 as negative class, determine the probability of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). No final calculation is required; only write the integration terms denoting the area under the curve for these probabilities.

(a) **(2 pts)** Probability of TP

(b) **(2 pts)** Probability of FP

(c) **(2 pts)** Probability of TN

(d) **(2 pts)** Probability of FN

3. **(MLE)** The probability density function for the Rayleigh distribution is given by: $f(x, \theta) = \frac{x}{\theta^2} exp(-\frac{x^2}{2\theta^2})$. Suppose Maria gathered data consisting of $X_1, X_2, ..., X_n$, which she deems are iid Rayleigh(θ) random variables and have the following likelihood function: $lik(\theta) = f(x_1, ..., x_n|\theta)$. Find the maximum likelihood estimate of Rayleigh's parameter (θ).

(a) **(2 pts)** Write the expression for the likelihood of θ.

(b) **(2 pts)** Write the expression for log likelihood of θ.

(c) **(2 pts)** Find out the maximum likelihood estimate for θ.

4. **(Naive Bayes)** (a) **(2 pts)** Consider a naive Bayes classifier with three boolean input variables, X_1, X_2 and X_3, and one boolean output, Y. Which probabilities must be estimated to train such a naive Bayes classifier? List them out. **Do not include probabilities that can be obtained with the constraint that the probabilities sum up to 1 for answering both (a) and (b).**

(a) **(4 pts)** How many probabilities would have to be estimated to learn the above classifier if we do not make the naive Bayes conditional independence assumption? List them out.

5. **(LDA) Hint:** (i) The outer product of two non-zero vectors always yields a rank=1 matrix. (ii) $rank(AB) \leq min\{rank(A), rank(B)\}$ (iii) Rank of sum of N rank-1 matrix is less or equal to N.

(i) **(1 pts)** Suppose that we have a set of n d-dimensional samples $\{x_1, ..., x_n\}$, n_1 in the subset $\mathcal{D}_2$ labeled ω_1 and n_2 in the subset $\mathcal{D}_2$ labeled ω_2. Write the expression for the sum of within-class scatter matrices S_w. Assume the mean of classes ω_1 and ω_2 are m_1 and m_2, respectively.

(ii) **(1 pts)** Now write the expression for between-class scatter matrix S_b.

(iii) **(1 pts)** Write the objective function whose maximum corresponds to the LDA projection vector.

(iv) **(2 pts)** Prove or disprove $rank(S_b) = 1$.

(v) **(2 pts)** Prove or disprove $rank(S_w) = 1$.

(vi) **(1 pts)** What will the direction of the vector $S_b w$ be for a $w \in R^d$? Why?

(vii) **(2 pts)** For multi-class cases, write down an expression for between-class scatter matrix using appropriate notations. If there are C classes, what will be the maximum possible rank of this matrix?

This page is intentionally kept blank, please do not forget to solve TRUE/FALSE questions in the next page. You may use this page for rough work.

6. **(10 pts)** Write TRUE or FALSE in capital letter and clear handwriting.

Write your roll number here:	

SN	Statement	TRUE/FALSE
1	If $a \in R^d$, then a trace of the matrix aa^t is equal to the Euclidean distance between the origin and the point a.	
2	PCA becomes equivalent to LDA if we project samples of each class using PCA independently.	
3	For a particular x, the value of the probability mass function can be greater than one.	
4	The PCA will project 2D-data points to $[1, 0]^T$ if $variance(x) > variance(y)$.	
5	If data distributions exactly follow y=x, then the rank of the covariance matrix will be zero.	
6	If X and Y are two real-valued random variables such that $Cov(X, Y) < 0$ then at least one of X or Y must have negative variance i.e. either $\sigma(X) < 0$ or $\sigma(Y) < 0$.	
7	Naive Bayes assumes that all features are conditionally independent given the class label.	
8	LDA is sensitive to outliers in the dataset.	
9	A die is loaded so that the probability of a face coming up is proportional to the number on that face. The die is rolled with outcome X. The expected value of X is greater than 4.	
10	$B = \{[1, 1]^T, [-1, 0]^T\}$ is one of the basis in 2D-space.	