MidTerm Exam
CS 331 / CS 530: Machine Learning
Course Instructor : Dr. Satyanath Bhat
Teaching Assistant : Rohit S, Akanksha Mishra, Anusree Debbarma
Date: 8 March 2022
Time: 14:00 to 16:00

INSTRUCTIONS:

- Answer all questions.
- Please upload your answers in pdf form.
- Define any notations as required.

1. Consider a two class problem with features in $\mathbb{R}$. The density of class-0 is normal with mean 0 and variance σ_1^2 and that of class-1 is normal with mean 0 and σ_2^2. The prior probabilities are $p_0 = p_1 = 0.5$. Find the Bayes Classifier (under 0-1 loss function) as a decision rule. (6 marks)

2. Show that sample mean ($\hat{\mu}$) is an unbiased estimator of the expectation (or population mean). Estimate the bias of the following estimators of variance

   (a) $\hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$.
   (b) $\hat{\sigma}_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$.

   (2+3+3 marks)

3. Consider a k class classifier. Let us suppose that the priors are given by $p_i = P(Y = c_i)$. Further suppose that class conditional densities are given by f_i. Prove that under 0-1 loss, the bayes classifier will chose label associated with the maximum posterior (10 marks)

4. Suppose (X, Y) are jointly distributed random variables with the following joint distribution

X	Y	P(X=x,Y=y)
1	1	0.3
1	2	0.1
3	1	0.3
2	1	0.15
3	2	0.15

   - What are the marginal distributions of X and Y?
   - What are the conditional distributions of X given Y and Y given X?
   - Consider this as regression problem and find the function $h(X)$ that minimizes the LMS error.
   - Model this as classification problem and find the classifier $h(X)$ that minimizes the 0-1 Loss. Find the probability of error for this classifier.

(3+3+7+7 marks)

5. Consider a model for regression where the target $y = w^T X + \epsilon$. Here ϵ is drawn independently from zero mean normal distribution with variance σ^2. Assume that $\{x_i, y_i\}_{i=1}^n$ are iid samples from this model. Derive the MLE estimate for the parameter w and σ^2 . (5 marks)

6. Consider a Bayesian model for regression where the target $y = w^T X + \epsilon$. The prior for the parameter w is standard normal. Here ϵ is drawn independently from zero mean normal distribution with known variance σ^2. Assume that $\{x_i, y_i\}_{i=1}^n$ are iid samples from this model. Derive the posterior distribution for the parameter w. (10 marks)

7. Write all the steps of the EM algorithm and briefly prove its convergence. (10 marks)

8. Consider two variants of logistic regression. Derive the gradient descent (ascent) update step for the following (relabelling is not permitted)

   (a) if the output label is $\{0, 1\}$

   (b) if the output label is $\{-1, 1\}$

   (4+4 marks)

9. A and V are square matrices of dimensions n and m respectively in the block matrix given below

$$P = \begin{bmatrix} A & U \\ V & C \end{bmatrix}$$

   (a) Given that A is invertible, obtain the inverse of the matrix P

   (b) Given that C is invertible, obtain the inverse of the matrix P

   (c) Derive the identity $(A - UC^{-1}V)^{-1} = A^{-1} + A^{-1}U(C - VA^{-1}U)^{-1}VA^{-1}$

   (6+6+4 marks)

10. Given that X_1, X_2, X_3 are jointly distributed as multivariate normal with mean $\mu = [\mu_1, \mu_2, \mu_3]^T$ and covariance Σ. Obtain the joint distribution of X_1, X_2. (7 marks)