IAI, TCG-CREST	October 15 2022

Machine Learning

Mid-Semester Examination

Ph.D. Program Session 2022-2023

Time Duration: 3 hours. *Full marks: 60.*

Answer any **6** of the following 10 questions.

1. Describe multiple linear regression and derive the update expressions for the model variables. Compare the methods of multiple linear regression and polynomial regression. (7+3)

2. Describe how and why the training and test errors change as the model complexity increases. How can residual plots be used to describe (i) non-linearity of variable-target relationships, and to (ii) identify outliers. (4+3+3)

3. What is the sigmoid function? Describe the Logistic Regression model. What are common choices of loss functions for the Logistic Regression model? Choose one loss function and derive expressions to update the model variables using gradient descent. (2+3+2+3)

4. Define a general Bayes Classifier that leads to an optimal decision rule that minimizes the overall risk.

   For a 2-class classification problem, let the class-conditional probabilities to each of the two classes w_1 and w_2 and the prior probabilities be defined as:

$$p(x|w_1) = \begin{cases} x & ;\, 0 < x < 1 \\ 2 - x & ;\, 1 \leq x < 2 \\ 0 & ;\, \text{o/w} \end{cases} \text{, with } P(w_1) = P \text{ , and,}$$

$$p(x|w_2) = \begin{cases} x - 1 & ;\, 1 < x < 2 \\ 3 - x & ;\, 2 \leq x < 3 \\ 0 & ;\, \text{o/w} \end{cases} \text{, with } P(w_2) = 1 - P.$$

   Assume that the two actions are defined in terms of classifying the data to each class, and the loss $\lambda_{11} = \lambda_{22} = 0$ and $\lambda_{12} = \lambda_{21} = 1$.

   1. Find the regions $\mathcal{R}_1$ and $\mathcal{R}_2$ where each class is decided by the Bayes Classifier. Also find the probability of misclassification for the Bayes Decision Rule.
   2. A classifier $\mathcal{C}'$ decides to classify points in region $\mathcal{R}_1 = (0, 1]$ to class 1, and decides to classify points in region $\mathcal{R}_2 = (1, 3)$ to class 2. Find the probability of misclassification for the decision rule of $\mathcal{C}'$.
   3. Show that the probability of misclassification by the decision rule of $\mathcal{C}' \geq$ the probability of misclassification by the Bayes Decision Rule.

   (3+4+2+1)

5. For a 3-class classification problem, where the actions are defined only in terms of classifying data to each of the 3 classes, derive Bayes Decision Rule and show that it minimizes the overall Risk. Also state the probability of misclassification of the Bayes Decision Rule.

   For binary classification using the Bayes classifier, let us assume that the data from both classes follow Gaussian distributions with equal σ^2 variance (i.e., all covariances are zero). Derive the equation of the decision boundary between the two classes. (6+4)

6. Compare discriminative and generative classification models, providing one example of each. Describe the maximum likelihood estimation procedure. Derive expressions for the maximum likelihood estimate of the parameters of a univariate Gaussian distribution. (4+2+4)

7. Define Conditional Independence, and explain how assuming Conditional Independence can reduce the number of distribution parameters that the Naive Bayes classifier needs to estimate. Assuming that the data from each dimension follow a Gaussian distribution, explain how the Naive Bayes classifier would estimate the relevant probabilities to classify the data. (6+4)

8. What is 'one-vs-all' and 'one-vs-one' classification, and discuss the drawbacks of these approaches. Describe a multi-class setup and prove that it overcomes the aforementioned limitations. Show how this setup can be followed to extend Logistic Regression for multi-class classification. (3+4+3)

9. Show that the assumption of symmetric and positive semi-definite matrices leads to describing a d-dimensional ellipsoid. Describe the problem setup for Principal Component Analysis. Derive the first principal component from the Principal Component Analysis objective function. (3+3+4)

10. Describe the k-Means clustering problem. How can the update expressions for the variables be derived from the k-Means problem. Describe Lloyd's alternating optimization algorithm for k-Means. (3+4+3)