Indian Institute of Engineering Science and Technology, Shibpur

B.Tech (CST) 7 th Semester End Semester Examinations, December 2021

Machine Learning(CS721/3)

Full Marks: 70   Time: 1 hr 30 minutes

Answer any five questions

1. a) Suppose the weights of randomly selected American female college students are normally distributed with unknown mean μ and standard deviation σ. A random sample of 10 American female college students yielded the following weights (in pounds):

115 122 130 127 149 160 152 138 149 180

Based on the definitions given above, identify the likelihood function and the maximum likelihood estimator of μ, the mean weight of all American female college students. Using the given sample, find a maximum likelihood estimate of μ as well.

   b) What is the trade-off between bias and variance?

   c) Describe Precision and Recall.   [7+4+3]

2. a) A teacher wants to decide whether a student can be given a pass mark, based on two features related to the attendance of the student, and his/her mid term marks. For simplicity, we model the two features with two binary variables $X1, X2 \in \{0,1\}$ and the class $Y \in \{0, 1\}$ where $Y = 1$ indicates that the customer can be given pass marks, and $Y = 0$ indicates otherwise.

   Consider the following dataset having four instances:
   $(X_1 = 0, X_2 = 0, Y = 0), (X_1 = 0, X_2 = 1, Y = 0)$

   $(X_1 = 1, X_2 = 0, Y = 0), (X_1 = 1, X2 = 1, Y = 1)$

   (i) Which model is better for the said application- logistic regression or linear regression? Explain briefly.

   (ii) Is there any logistic regression classifier using $X1$ and $X2$ as features, that can perfectly classify the given data?

   (iii) If we change the first instance to $(X_1 = 0, X2 = 0, Y = 1)$ can there be any logistic regression classifier using X_1 and $X2$ as features, that perfectly classifies the data?

   b) What are the characteristics of convolution neural network (CNN) in comparison with fully connected Feedforward neural network?   [9+5]

3. a) Explain the significance of the RELU activation function in Convolution Neural Network(CNN).

   b) What are the different types of Pooling? Explain their characteristics.

   c) Why do we prefer Convolutional Neural networks (CNN) over Artificial Neural networks (ANN) for image data as input?   [5+5+4]

4. a) Describe estimation and prediction of multiple regression. How is linear regression different from multiple regression?

b) For the following random samples, find the likelihood function:

  i. Xi ~Binomial(3,θ) and we have observed
(x1,x2,x3,x4)=(1,3,2,2)(x1,x2,x3,x4)=(1,3,2,2).

ii. Xi ~Exponential(θ) and we have observed
(x1,x2,x3,x4)=(1.23,3.32,1.98,2.12)(x1,x2,x3,x4)=(1.23,3.32,1.98,2.12).   [7+7]

5. a) What is the objective of an Auto Encoder?

   b)Explain functionality of Auto Encoder in support of your answer of question 5.a.

   c) What is the basic idea behind variational auto encoder?   [5+5+4]

6. a) Derive a gradient descent training algorithm that minimizes the sum of squared error cost function, for the following hypothesis:
$$h_B(x) = B_o + B_1x_1 + B_1x_1^2 + B_2x_2 + B_2x_2^2 + ... + B_nx_n + B_nx_n^2$$

where $(x_1, x_2, ..., X_n)$ represents an instance having n features and $B_i, 0 <= i <= n$ represent the parameters to be learned. Assume that the number of training instances is m. Give your answer in the form $B_j := B_j + ...$ for $1 <= j <= n$.

   b) Consider a single LSTM cell whose inputs are the current input x_t and h_{t-1} and c_{t-1} from the previous step; and the outputs are c_t and h_t. Assume that the input dimension is D and the dimension of h_t is H. Sketch a LSTM cell showing the various gates inside the cell, including forget gate f_t , input gate (i_t), o_t , c_t , h_t.

   c) What are the drawbacks of RNNs?   [6+4+4]