JK Lakshmipat University, Jaipur
Institute of Engineering and Technology
**Machine Learning (CS1138)**
Practice Questions II, BTech, IV Semester

Instructor: Dr. Arpan Gupta Date: 15 - 03 - 2025

**Practice Questions. Some questions have already been discussed in class**

1. Using logistic regression, show the gradient update step and find the updated values of $\Theta$, for 1 iteration on the following data. The initial values of $\Theta_0 = \Theta_1 = \Theta_2 = 0$, Learning Rate ($\alpha = 0.1$). Also, calculate the cost ($J(\Theta)$) for initial and updated values of $\Theta$.

| $x_1$ | $x_2$ | **Y** |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 3 | 3 |
| 3 | 4 | 3 |

2. What do you mean by a convex function? Define mathematically. If a cost function is non-convex, what method can be used in order to increase the chances of reaching the global optima when applying gradient descent.

3. What do you mean by Bias-Variance Tradeoff. Show the graphs for high bias and high variance cases.

4. Consider a set of five training examples given as $((x_i, y_i), c_i)$ values, where $x_i$ and $y_i$ are the two attribute values (positive integers) and $c_i$ is the binary class label:

   $\{((1,1),-1),\ ((1,7),+1),\ ((3,3),+1),\ ((5,4),-1),\ ((2,5),-1)\}$.

   Classify a test example at coordinates $(3,6)$ using a k-NN classifier with $k = 3$ and Manhattan distance. Your answer should be either +1 or -1. Clearly show the steps.

5. What is the difference between euclidean and non-euclidean distance measures. Mention some non-euclidean distance measures and their applications. Which type of distance measures follow the triangle inequality?

6. What is the best way to choose the value of k, in k nearest neighbours? Explain.

7. Suppose we want to compute 10 - Fold Cross Validation error on 100 training examples. We need to compute error N1 times, and the cross validation error is the average of the errors. To compute each error, we need to build a model on the data of size N2 and test the model on data of size N3. What are the appropriate numbers for N1, N2, N3?

   1. N1 = 10, N2 = 100, N3 = 100
   2. N1 = 1, N2 = 90, N3 = 10
   3. N1 = 10, N2 = 90, N3 = 10
   4. N1 = 10, N2 = 100, N3 = 10

8. Consider the following vectors A and B in a 3D space, compute their cosine similarity. Also, mention some applications/tasks where cosine similarity is used.

   A = [1, 2, 3] B = [4, 5, 6]

9. A logistic regression model generates the following probability values $h_\theta(x)$ for the binary classification task, where $\mathbf{Y}$ are the ground truth labels.

| $\mathbf{Y}$ | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|
| $h_\theta(x)$ | 0.14 | 0.83 | 0.42 | 0.71 | 0.31 | 0.26 | 0.75 | 0.92 | 0.55 | 0.64 |

   (a) Find the precision, recall, and F1 scores for the above model assuming threshold value for $h_\theta(x)$ is 0.7.

   (b) Draw the ROC curve for the model by taking FPR on the x-axis and TPR on the y-axis. Take the threshold values for $h_\theta(x)$ at [0, 0.2, 0.4, 0.6, 0.8, 1].

   (c) Draw the curve (on the same plot) that denotes a "random" prediction model. Which of the two is a better model.

10. A neural network takes $X \in \mathbb{R}^4$ as input and consists of 2 hidden layers with sizes {5, 4} and finally generates an output for a multi-class classification problem with $y \in \mathbb{R}^3$.

   (a) Draw the complete network (with edges and bias nodes) and label the input, hidden, and output layers along with their dimensions.

   (b) Write the forward propagation equations in a vectorized form, along with the dimensions of the vectors/matrices involved.

   (c) Calculate the total number of learnable parameters/weights in the model.

11. Given a data set $D = \{o_1, ..., o_n\}$ with known class labels $Y(o_i) \in C = \{A, B, C\}$ of the objects. In order to evaluate the quality of a classifier $K$, each object $o_i \in D$ is additionally classified using $K$, yielding class label $K(o_i)$. The results are given in the table below.

| $i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y(o_i)$ | A | B | A | C | C | B | A | A | A | B | B | C | C | C | B |
| $K(o_i)$ | A | A | C | C | B | B | A | A | A | C | A | A | C | C | B |

   (a) Setup the confusion matrix.

   (b) Compute the accuracy / classification error.

   (c) For each class $i \in C$ compute the precision and recall values.

   (d) Compute the F1 score.

12. (Advanced) "A k-nearest neighbors (k-NN) classifier's error is at most twice that of the optimal Bayes classifier (as the number of training samples increases)". Using which theorem can we prove this statement. Explain. Refer `https://www.cs.cornell.edu/courses/cs4780/2018fa/lectures/lecturenote02_kNN.html`.

13. (Advanced) Write down the Perceptron learning algorithm and discuss its convergence. Refer `https://www.cs.cornell.edu/courses/cs4780/2018fa/lectures/lecturenote03.html`