AI42001 Spring 2023 **Lab Test 1** **9th March 2023**

Time: 2:15 PM -- 3:45 PM **Total marks: 20**

Use the last 5 digits of your roll number as the seed of the random number generator.

Q1. A) Choose a random 5D vector W, where the first dimension lies between (0,1), 2nd dimension between (1,2), 3rd between (2,3) etc. **[1 mark]**

B) Use it to generate N=1000 data-points $\{X_i, Y_i\}$, where $Y_i = W.X_i + e_i$, where X_i is a 5D feature vector where each feature is uniformly between (-10, 10) and e_i is a random value between -1 and 1.

[1 mark]

C) Fit a linear regression model on (Xi, Yi) and estimate W1. **[2 marks]**

D) Calculate the estimation error ||W-W1||2 and the MSE $(1/N)*\sum(Y_i-W1.X_i)2$. **[2 marks]**

E) Fit a LASSO regression model on (X_i, Y_i) and estimate W2 for different values of λ (1 to 20). **[2 marks]**

F) Sparsify each W2 to obtain W3 by setting all values below a certain threshold to 0 (you can choose the threshold). Plot the MSE $(1/N)*\sum(Y_i-W3.X_i)2$ as a function of λ. **[2 marks]**

You can use library functions for Linear and LASSO regression.

Q2. A) Choose a random 10D vector W, where each value lies between -1 and 1. Generate N=100 datapoints (X_i, Y_i) such that X_i is a 10D vector whose elements lie between either (-10,-1) or (1,10), and $Y_i=\text{sign}(W.X_i)$. Basically you create a linearly separable dataset with wide margin. **[2 marks]**

B) Implement perceptron algorithm to find a separating hyperplane. **You should not use library functions in this case.** **[3 marks]**

C) Plot the number of updates you make in each iteration. **[1 mark]**

D) Now, reduce the margin of the dataset by adding 10 points, whose features lie between between (-1,0) and (0,1). Labels are still generated as $Y_i=\text{sign}(W.X_i)$, i.e. the dataset is still linearly separable but now has narrower margin. **[1 mark]**

E) Run the perceptron algorithm again and plot the number of updates in each iteration. **[1 mark]**

F) Now add 5 "outliers" on both sides of W. **[1 mark]**

G) Run the perceptron and plot the number of updates in each iteration. **[1 mark]**