Indian Institute of Information Technology Guwahati
Machine Learning Lab (CS360)
End Semester Examination

The exam is only for CG1 students

Instructions:

1. Select your question (only one question out of five) using the following formula

   Question number = Your Roll Number + Your day of birth) % 5) + 1
   Example:
   Say your date of birth is 13-Dec-2001 and roll number is 1801002
   Question number = 2 + 13) % 5) + 1 = 1

2. Please share your google colab link using the google form

3. Export the Colab code to a pdf after executing the code and send the pdf (File -> Print -> Save as PDF)

4. Fillup the following form:
   https://forms.gle/U1uAMzmV8FwZ8Cd89

5. Exam time: 3 hours, please submit by 12:30 (16-Nov-2020)

Question 1: Implement a Fuzzy version of a self-organizing feature map (SOM) by combining SOM with Fuzzy c-means (FCM) on the Iris dataset.

1. Do usual SOM
2. Apply FCM on the cluster centers following the Equations 1-5 in Page 2-3 from [1].

(Please find the reference paper is in the attachment. File name 1.pdf)

Question 2: Implement a modified version of the *k*-means algorithm [4] for the Iris datasets. To do the modification you need to follow the "**algorithm A**" part only of [4].

(Please find the reference paper is in the attachment. File name 4.pdf)

Question 3: Implementation of PCA and MLP

a) Train MLP with Pavia University hyperspectral image datasets. Take all 103 numbers of features and optimize the model. Here, optimization means the identification of the optimal learning rate and hidden neuron.

b) Now reduce the feature dimension from 103 to some optimal number using PCA. The new dimension is less than the actual feature dimension. Finally, train MLP with the same optimal parameters (as implemented in (a)). The input for MLP will be the reduced feature dimension.

c) Report a graph by varying the number of feature dimensions (103:20:23) against accuracy.

Dataset download: [Click here](#)

Convert matlab file to numpy: [Click here](#) and [Click here](#)

PCA implementation resources: [Click here](#) or [Click here](#)

Question 4: Implements the linear regression using a Million Song Dataset from the UCI Machine Learning Repository. Your goal is to train a linear regression model to predict the release year of a song given a set of audio features.

a) Take a subset of data randomly from the dataset to create the linear regression model.

b) Next task is to extract a subset of features from the dataset using the PCA techniques. Thereafter you can use those features into the regression model.

Dataset link: [Click here](#)

Question 5:

a) Write a program to implement the KNN classifier for a IRIS dataset. Compute the accuracy of the classifier, considering few test data sets.

b) Write a program to implement the k-means and DB scan considering the same IRIS dataset. For k-means take 3 numbers of clusters.

c) In the k-means implementation measure the accuracy by identifying the each misclassified samples in the clusters. Assume that each cluster will represent a class.