IE406: Machine Learning Lab

Assignment 4

(Date: 19/09/2021)

1. The MNIST database of handwritten digits, has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger set available from NIST. It is a good database for people who want to try learning techniques and pattern recognition methods on real-world data while spending minimal efforts on pre-processing and formatting.
   (hint : use scikit-learn library's "fetch_mldata" to load dataset)

   Plot **Mean Image** of all the 10 digits.

2. Perform Linear Discriminant Analysis (LDA) on the MNIST dataset* for binary as well as for multiclass classification. Plot confusion matric and find out the combinations where the classifier is confused in predicting the right label.

3. Perform Quadratic Discriminant Analysis (QDA) on the MNIST dataset* for multiclass classification. Plot confusion matric and find out the combinations where the classifier is confused in predicting the right label.

4. Perform Naïve-Bayes on the MNIST dataset* for multiclass classification. Plot confusion matric and find out the combinations where the classifier is confused in predicting the right label.

5. Mean and variance of two classes are,

   Class_1 : $\mu = 8$, $\sigma^2 = 20$
   Class_2 : $\mu = 16$, $\sigma^2 = 25$

   a. Draw 50 random samples from N[5,20]

   b. Draw 50 random samples from N[11,10].

   c. Draw 50 random samples from N[20,8]

   and classify using Naïve-Bayes classifier having apriory probabilities as (0.5,0.5), (0.3,0.7) and (0.7,0.3) and visualize data and class by plotting histogram.

Note:

***Keep Train:Test dataset ratio as 9:1.**
Submission Deadline: 11:59 PM, Saturday, 2nd October 2021
Strictly follow the submission guidelines given in the classroom.