Description

Download the MNIST dataset from here. Pick its training and test sets. Each sample in the dataset is represented by a 784-dimensional vector. There are ten classes (the ten digits, `0' to `9'), and each sample is associated with one class.

(1) Using the raw binary features, implement the streaming Naive Bayes algorithm and classify the test data.
(2) Project the binary features into a lower dimensional space using a dimensionality reduction technique of your choice (such as PCA, LDA, t-SNE, etc.), by varying the dimensionality in the range {50,100,200}, and classify the test data using the same algorithm.
(3) Compare (1) and (2) in terms of classification accuracy and time required during the training and testing phase.
(4) Compare (1) and (2) with the classification accuracy you had obtained in assignment-2.

Deliverables

(1) A folder containing your codes and a detailed readme file. You may use any programming language.
(2) A report (PDF) describing the experimental details, results, analyses, observations, etc.
(3) Create a single zipped file name <RollNo_Assig3.zip> containing the above two and upload.

General instructions

(1) Do not paste your codes in the report.
(2) Cite all the resources in the report.
(3) If anything is missing or not clear from the above description, you may make appropriate assumptions and clearly mention them in the report.
(4) A submission which does follow any of the guidelines will be awarded a penalty.
(5) Any submission received after the deadline will be awarded a penalty. The time recorded in google-classroom will be considered.
(6) Plagiarism of any kind will result in a zero in this assignment, and an additional penalty in the total score in the course.