Ahsanullah University of Science and Technology
Department of Computer Science and Engineering (CSE)

CSE 4108: Artificial Intelligence Lab, Spring 2018

Lab Group: All Offline: 6 Topic: Naive Bayes&Logistic Regression

---

Task 1. You need to apply Naive Bayes algorithm and logistic regression (using stochastic gradient descent) to classify the given dataset "banknote.csv".
"banknote.csv" dataset contains 1372 rows, each with 4 features and one target value (0(forgery) /1(authentic)). You can use pandas library for reading from CSV files.
What you need to do are:

1) Read the dataset
2) Divide the dataset in 80%-20% proportion after shuffling it. Keep 80% of rows as training dataset and rest of 20% dataset as test dataset.
3) Apply Naive Bayes on training dataset to learn prior, posterior and likelihood probabilities. Remember, as all the features are continuous, you need to apply Gaussian Naive Bayes.
4) Apply your Naive Bayes model on test dataset to predict their target values (0/1) and compare them with the real values.
5) Apply logistic regression (stochastic gradient descent) on training dataset to learn the parameters (weights/θ).
6) Apply your regression model on test dataset to predict their target values (0/1) and compare them with the real values.
7) For each algorithm working on test dataset, count these values:
    a) How many 0's were predicted as 0? (True Negative)
    b) How many 0's were not predicted as 0? (False Positive)
    c) How many 1's were predicted as 1? (True Positive)
    d) How many 1's were not predicted as 1? (False Negative)
8) Remember, you must do this assignment in Python. Write the algorithms from scratch. No plotting is needed as shown in sample code.