DEEP LEARNING ACADEMY

Today we are going to implement the Naive Bayes classifier in python language and test it on the Pima Indians Diabetes dataset. While you are free to make your own implementation, it is recommended to follow the implementation steps below and test each one to make sure it works properly.

1. **Handle Data**: Load the data from CSV file and split it into training and test datasets. The dataset is available at: https://www.kaggle.com/uciml/pima-indians-diabetes-database
   Suggestion – first understand your data (this is a sufficient summary - https://www.andreagrandi.it/2018/04/14/machine-learning-pima-indians-diabetes/) by exploring it's columns, values, possible pitfalls, etc.
2. **Summarize Data (train)**: summarize the properties in the training dataset by calculate for every feature and class (prediction value) the mean and the std.
3. **Write a function which make a prediction**: Use the summaries of the dataset to generate a single prediction, which based on the gaussian distribution with the corresponding mean and std of each of the features. You can find the equation for the probability of an event given a Gaussian distribution in: https://en.wikipedia.org/wiki/Naive_Bayes_classifier#Gaussian_naive_Bayes
4. **Make Predictions**: Generate predictions on the whole test dataset.
5. **Evaluate Accuracy**: Evaluate the accuracy of predictions made for a test dataset as the percentage correct out of all predictions made.
6. **Tie it Together**: Use all of the code elements to present a complete and standalone implementation of the Naive Bayes algorithm.
   * (Optional) Try building it into a class with fit(train) method which calculates the mean and std and predict(test) method which makes a Naive Bayes prediction for the test data.

We are going along the instructions from the following link:

http://machinelearningmastery.com/naive-bayes-classifier-scratch-python/

Part 2: Classes (Optional part):

For tutorials refer to any of the internet tutorials on python classes, such as:
https://www.learnpython.org/en/Classes_and_Objects
https://en.wikibooks.org/wiki/A_Beginner%27s_Python_Tutorial/Classes
https://docs.python.org/3/tutorial/classes.html#classes

Rewrite your code to use class such that it contains:

1. A DataSet class
   a. It should be instantiated (__init___ function) with a dataset (with labels)
   b. It should contain a function which gets a percentage and returns the data, after randomly permuting it, split to train and test according to that percentage.
2. A NaiveBayes class
   a. Contains function which classifies test data
   b. the class should also contain all the relevant functions for the classification calculation.