Naive-Bayesian-Classifier

Builds a naive bayesian clasifier to predict Response using dataset-C

Table of Contents

Requirements

Running this script requires python3+ and pandas,math,random,copy,sklearn,matplotlib libraries of python3 to be installed on running machine

Setup

Run on command line

```
sudo apt-get update
sudo apt-get install python3.6
pip3 install pandas
pip3 install random
pip3 install math
pip3 install copy
pip3 install sklearn
pip3 install matplotlib
```

This will install all required libraries to run this script on the machine
If few of the libraries are already installed,run only the required commands from above

Running the Program

Run on command line

```
python3 script.py
```

By default,the program executes all three parts asked in question

Part-1

Program first loads the data into pandas array and encodes string data into numerical class values and trains a naive bayesian classifier using five fold cross validation and prints corresponding accuracies

Part-2

Applies PCA function to reduce dimensions
Prints PCA components vs variance graph
Trains a naive bayesian classifier using five fold cross validation and prints corresponding accuracies

Part-3

Removes Outlier Features
Runs sequential backward selection method to reduce dimensions
Prints final set of attributes retained
Trains a naive bayesian classifier using five fold cross validation and prints corresponding accuracies

Running each part seperately

Open script.py provided in folder, go to the main function(at the bottom of the file) and comment out corresponding lines to check results of only individual parts.
(Each part is described in comments in the file)
Part-1 : Comment out Parts 2-3
Part-2 : Comment out Part-1 and Parts 3
Part-3 : Comment out Part-1-2