Due Date: Sunday 22nd April (Midnight)

Naïve Bayes Classifier

CS 331 Assignment 3

Instructor: Dr. Yasir Mehmood

**Assignment 3**

**Naïve Bayes Classifier**

**Honor Code**

You are to do this assignment by yourself. All of the submitted code should be your

Creation and a result of your own thought process. No cooperation is allowed under any circumstances. Any help outside of the course staff is prohibited. You are also given the responsibility of reporting any cooperation incidences to the course staff. All code will be checked for similarities, and cases will be promptly forwarded to relevant authorities for action.

**There will be no extensions, everything is clearly explained in this handout along with detailed notes, formulas, explanation of data set and problem and what is expected of you. Please start early in order to avoid facing problems on the last day and submitting incomplete assignments. Do your best and ask for help from the course staff when stuck.**

**Important Instructions:**

* **You are not allowed to use any pre-existing machine learning libraries. You need to implement the naïve Bayes algorithm yourself.**
* This assignment needs to be done in Python.
* You are required to use Numpy where necessary.
* You have to take command line arguments:

**python classifier.py Spect\_train Spect\_test**

Make sure you follow this format.

**Naïve Bayes Classifier**

The goal of this assignment is to implement Naïve Bayes Classifier. You will use the Bernoulli naïve Bayes model for the classification task. Bernoulli model requires that all attributes value is binary as a result the dataset of SPECT, provided to you, contains only binary values.  
  
**Explanation of dataset:**

Each patient is classified into two categories: Normal and Abnormal, depending on the number of medical tests he/she passes. The database contains 267 patients’ data, every person underwent 22 medical tests and each test was either pass or fail. As a result, for each patient 22 binary values were extracted.  
  
You have been provided with two files, Spect\_train and Spect\_test. Spect\_train has a total 80 data points and Spect\_test has 187 data points.

You will use Spect\_train patient data to train your naïve Bayes classifier and Spect\_test to test it.

A single patient in the dataset is described as a single line of the file. So each line has 23 values, the first value of each line describes whether the person was described as normal (value of 1) or abnormal (value of 0). All other 22 values define which test number the patient failed and which he/she passed.

For example:

The first line of SPECT\_train is 1,0,0,0,1,0,0,0,1,1,0,0,0,1,1,0,0,0,0,0,0,0,0.

It can be interpreted as:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Patient | Test1 | Test2 | Test3 | Test4 | Test5 | … | Test20 | Test21 | Test22 |
| Normal (1) | Fail (0) | Fail (0) | Fail (0) | Pass (1) | Fail (0) | … | Fail (0) | Fail (0) | Fail (0) |

If the above data point started from zero: 0,0,0,0,1,0,0,0,1,1,0,0,0,1,1,0,0,0,0,0,0,0,0

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Patient | Test1 | Test2 | Test3 | Test4 | Test5 | … | Test20 | Test21 | Test22 |
| Abnormal (0) | Fail (0) | Fail (0) | Fail (0) | Pass (1) | Fail (0) | … | Fail (0) | Fail (0) | Fail (0) |

**Task**:

You have to implement the Bernoulli naïve Bayes classifier for the above set such that given 22 medical test reports of a person, your classifier predicts whether the person is normal or abnormal. You will test your classifier using Spect\_test file.

You also have to write a short report on the process and you will be marked on the quantity of the content, so write anything you feel deserves credit. (Should not be more than a page). Take both Spect\_test and Spect\_train files as input from command line as mentioned in the instructions. Both your training and testing code will be in the same file (classifier.py), you can divide it in classes or functions as appropriate.

**Sample Output:**

You should correctly show your output on the screen.

##########

*Starting to Train on 80 data points . . .*

*Training Complete*

*Testing on 187 data point* . . .

*Total Accuracy := 90%*

##########

**Marking Criteria:**

Implementation 45

Accuracy 30

Report 20

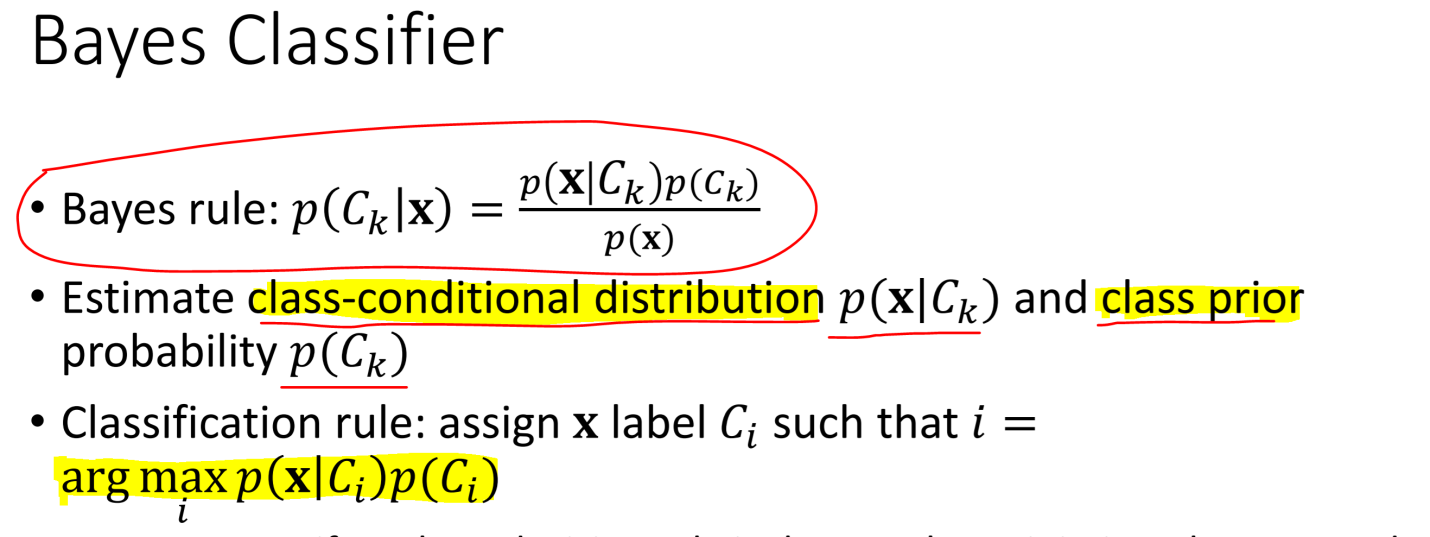
Output 5

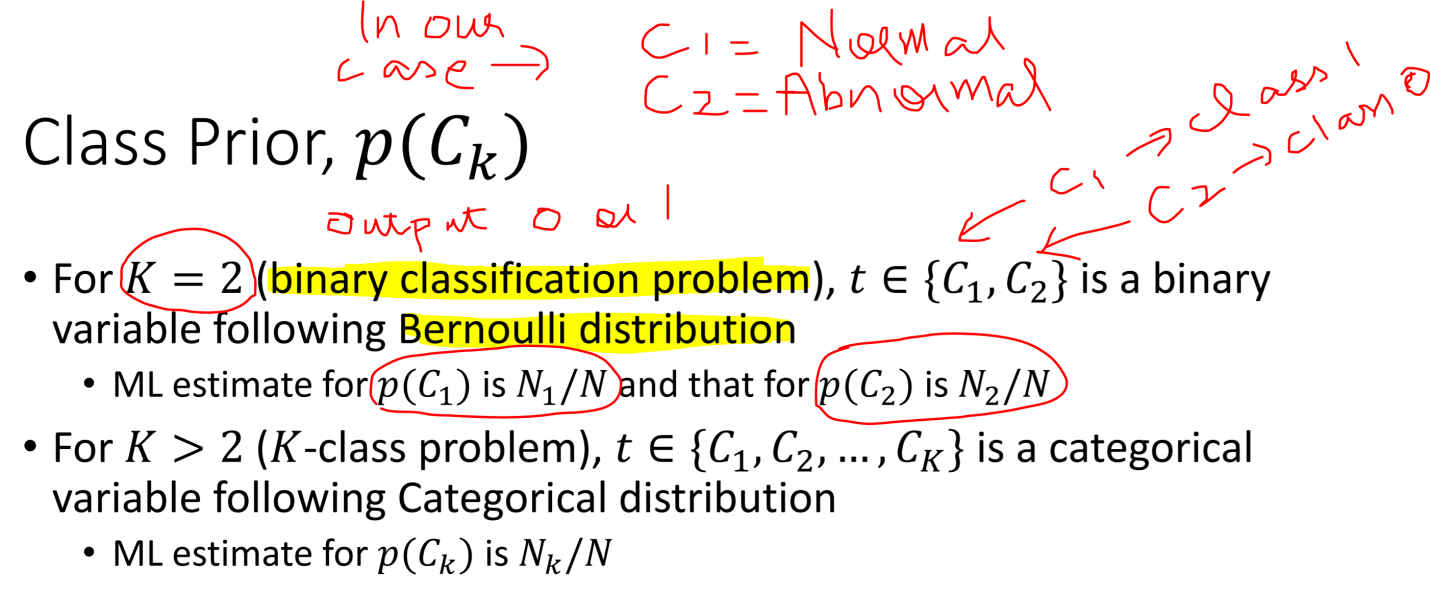
**Files to Submit:**

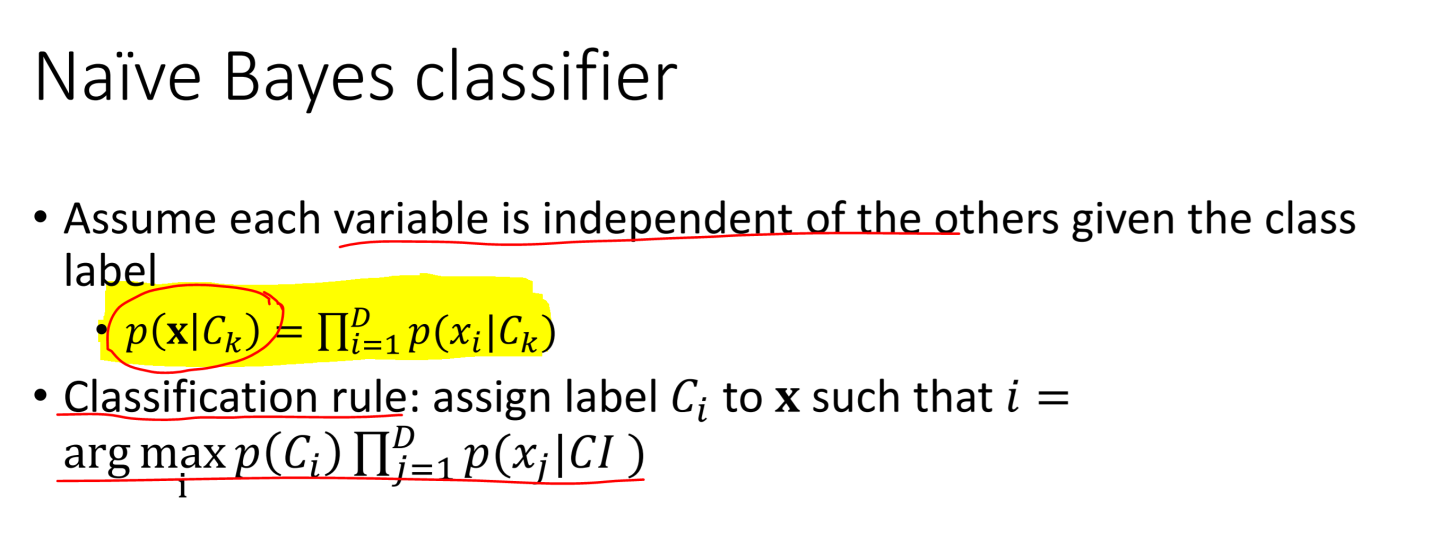
* Classifier.py
* Report.txt

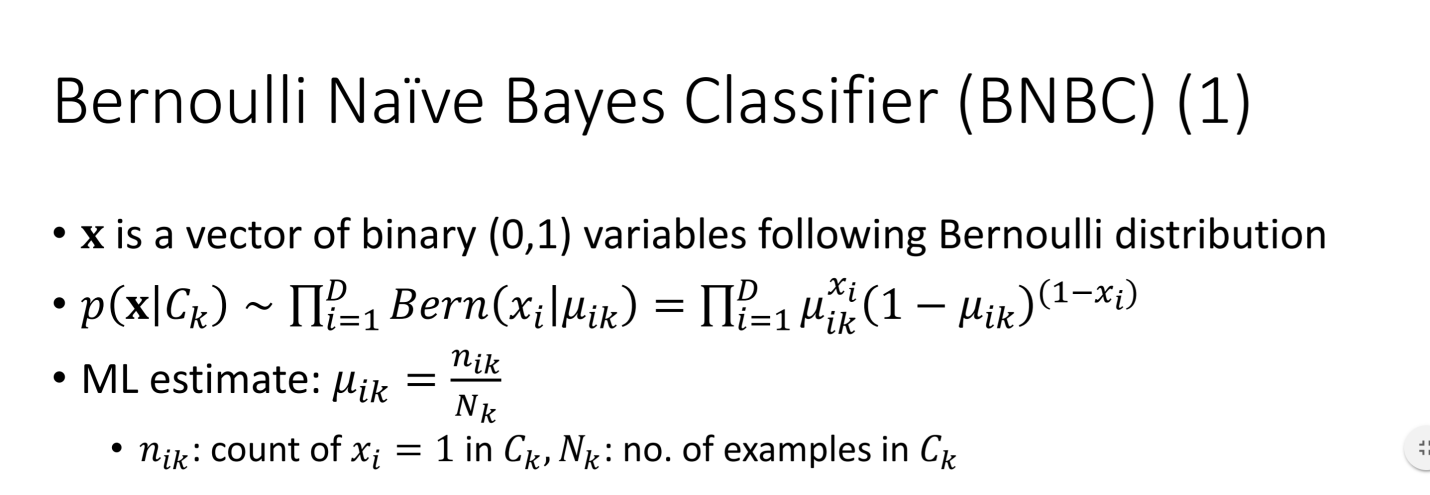
Zip all the files in a folder named: **YourRollNumber\_Assignment3.zip**. All files will go through Moss and no plagiarism act will be tolerated.

Some Notes below to help you Get Started:









Start Early and Good Luck! ☺