In this homework, you will implement a naïve Bayes' classifier in R, Matlab, or Python. Here are the steps you need to follow:

1. Read Section 5.7 from the textbook.

2. You are given a multivariate classification data set, which contains 195 handwritten letters of size 20 pixels × 16 pixels (i.e., 320 pixels). These images are from five distinct classes, namely, A, B, C, D, and E, where we have 39 data points from each class. The figure below shows five sample figures from each class. You are given two data files:
   a. `hw01_data_set_images.csv`: letter images,
   b. `hw01_data_set_labels.csv`: corresponding class labels.
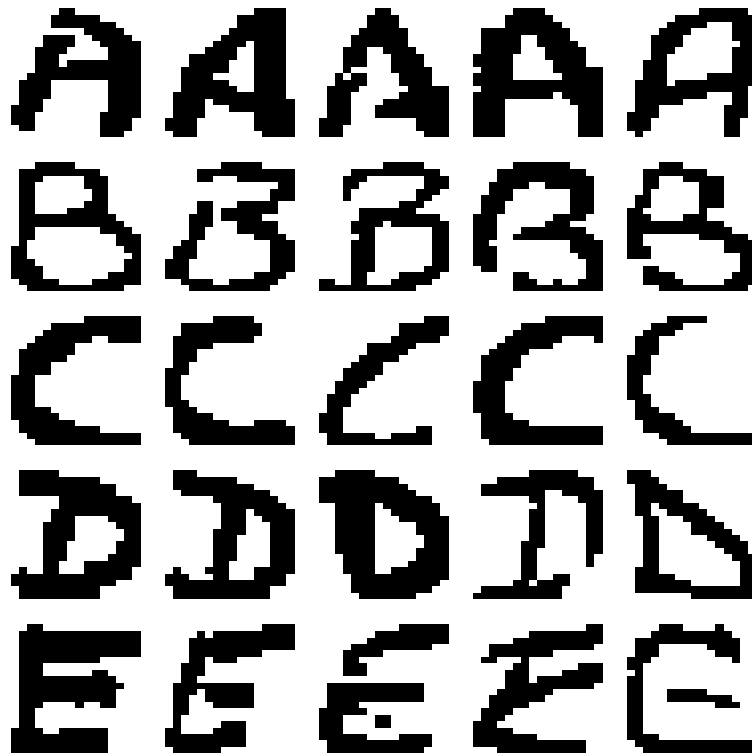
3. Divide the data set into two parts by assigning the first 25 images from each class to the training set and the remaining 14 images to the test set.

4. Estimate the parameters $\widehat{p}_1, \widehat{p}_2, \widehat{p}_3, \widehat{p}_4, \widehat{p}_5, \widehat{P}(y = 1), \widehat{P}(y = 2), \widehat{P}(y = 3), \widehat{P}(y = 4)$, and $\widehat{P}(y = 5)$ using the data points you assigned to the training set in the previous step. Your parameter estimations should be similar to the following figures.

```
> print(pcd[,1])
[1] 0.00 0.00 0.00 0.04 0.04 0.04 0.16 0.20 ...
> print(pcd[,2])
[1] 0.04 0.24 0.24 0.20 0.12 0.08 0.12 0.16 ...
> print(pcd[,3])
[1] 0.00 0.00 0.00 0.00 0.00 0.12 0.20 0.24 ...
> print(pcd[,4])
[1] 0.12 0.44 0.40 0.16 0.12 0.08 0.08 0.08 ...
> print(pcd[,5])
[1] 0.00 0.12 0.12 0.08 0.12 0.16 0.12 0.04 ...
```

5. Calculate the confusion matrix for the data points in your training set using the parametric classification rule you will develop using the estimated. Your confusion matrix should be similar to the following matrix.

```
        y_train
y_hat   1   2   3   4   5
    1  25   0   0   0   0
    2   0  24   1   0   1
    3   0   0  24   0   0
    4   0   1   0  25   0
    5   0   0   0   0  24
```

6. Calculate the confusion matrix for the data points in your test set using the parametric classification rule you will develop using the estimated parameters. Your confusion matrix should be similar to the following matrix.

```
        y_test
y_hat   1   2   3   4   5
    1   7   0   0   0   0
    2   0  11   3   2   4
    3   0   0   7   0   0
    4   7   3   3  12   0
    5   0   0   1   0  10
```

What to submit: You need to submit your source code in a single file (.R file if you are using R, .m file if you are using Matlab, or .py file if you are using Python) and a short report explaining your approach (.doc, .docx, or .pdf file). You will put these two files in a single zip file named as *STUDENTID.zip*, where *STUDENTID* should be replaced with your 7-digit student number.

How to submit: E-mail the zip file you created to vbakir@ku.edu.tr with the subject line *Intro2MachineLearningHW01*. Please follow the exact style mentioned for the subject line and do not send a zip file named as *STUDENTID.zip*. Submissions that do not follow these guidelines will not be graded.

Late submission policy: Late submissions will not be graded.

Cheating policy: Very similar submissions will not be graded.