**Machine Learning**

**Fall 2022**

**HW7**

**Due: Nov. 6 (Sunday) 11:59pm, via Blackboard**

**Problem 1**

In this exercise. you are required to make prediction of the test sample (a melon) if it is good/ready for consumption using the Naïve Bayes Classifier.

In this dataset (17 samples), each row corresponds to a sample. The first four columns are four features and the last column is the label/class.

**Table 1 (17 samples)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Color** | **Texture** | **Root** | **Touch** | **Good melon (label)** |
| Pitch black | clear | curled | Hard and smooth | yes |
| dark green | clear | curled | Hard and smooth | yes |
| Light white | clear | curled | Hard and smooth | yes |
| dark green | clear | slightly curled | Soft sticky | yes |
| Pitch black | slightly fuzzy | slightly curled | Soft sticky | yes |
| Pitch black | clear | slightly curled | Hard and smooth | yes |
| Pitch black | slightly fuzzy | slightly curled | Hard and smooth | no |
| dark green | clear | straight | Soft sticky | no |
| Light white | fuzzy | straight | Hard and smooth | no |
| Light white | fuzzy | curled | Soft sticky | no |
| dark green | slightly fuzzy | slightly curled | Hard and smooth | no |
| Light white | slightly fuzzy | slightly curled | Hard and smooth | no |
| Light white | fuzzy | curled | Hard and smooth | no |
| dark green | slightly fuzzy | curled | Hard and smooth | no |

Please determine whether the melon (dark green, slightly fuzzy, curled, soft sticky) is good or not.

**Problem 2**

In this exercise. you are required to write a Matlab code of a Naïve Bayes model to classify the dataset ‘**dataset\_for\_naive\_bayes.csv**’.

**Table 1 (0:class 1, 1: class 2)**

|  |  |  |
| --- | --- | --- |
| **Exam 1** | **Exam 2** | **Status** |
| -0.1589 | 0.42397 | 1 |
| -0.3479 | 0.4707 | 1 |

**.**

**.**

**.**

|  |  |  |
| --- | --- | --- |
| -0.31694 | -0.22895 | 1 |
| -0.29412 | -0.1348 | 0 |
| -0.15311 | 0.1845 | 0 |

Please divide the entire dataset into two subsets (training and test sets) in a 7:3 ratio, which means 70% of the dataset is the training set and the remaining 30% is the test set. Based on the training set, you are required to train a Naïve Bayes classifier and implement the trained Naïve Bayes classifier on the test set.

1. Plot the original dataset.
2. Draw the decision boundary and calculate the accuracy. Include the training set and it’s decision boundary in a figure, then calculate the accuracy on this training set. In another figure, include test set and it’s decision boundary, then calculate the accuracy on this test set.

**Note**: When you submit the homework, you should make sure that the homework includes following items:

**a matlab file:**

maincode.m / maincode.mlx

**A report** that briefly addresses

1) accuracy on training and test sets.

2) decision boundary figures.

3) a flow chart of your codes.

4) all other results that you think are necessary.

If you make changes to the dataset file, such as changing and adding the column name, please upload your dataset to BB.

If we can run the code neither under your modified dataset nor under the dataset posted online, you may get low grade

**Hints:**

1. Since the data has already been normalized, you do not need to preprocess the dataset.
2. Make sure your accuracy on the training set and test set>=90%.