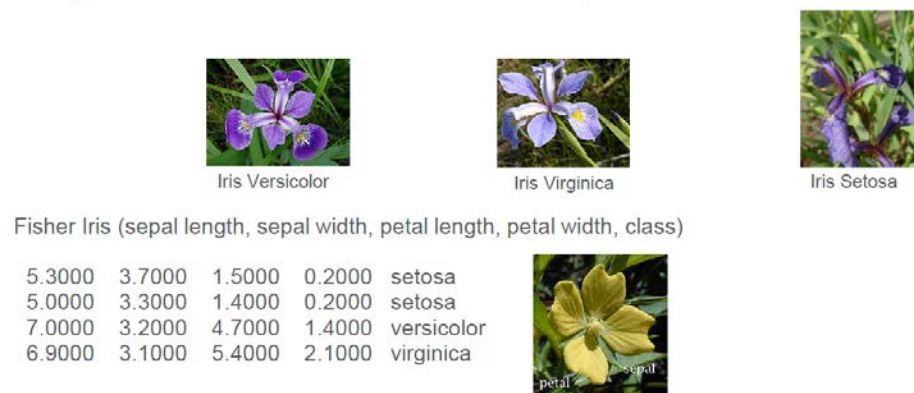


[50 points]

Iris flower dataset (iris_dataset.dat)

- Introduced by Ronald Fisher in 1936 (sometimes called Anderson's Iris dataset)
- The first four columns list the measurements in centimeters of the variables sepal length, sepal width, petal length and petal width.
- The last column shows the class label of each example (1=setosa, 2=versicolor, 3=virginica)



In this coding assignment you are to implement a Minimum Risk Bayes Decision Theoretic classifier and use it to classify the test examples in the provided datasets.

Assume the following:

- All conditional density functions are multivariate Gaussian
- Each class has its own covariance matrix
- Equally likely prior probabilities
- 0-1 loss function
- 10-fold cross validation for training & testing

The suggested MATLAB commands are included. You don't have to use these commands. Type "doc <command>" at the MATLAB prompt to see more information on the commands.

1. Read "iris_dataset.dat" and randomly shuffle the data to avoid class imbalance
2. Using the training data, estimate the parameters by MLE method
 - MATLAB commands: mean, inv
3. Classify test data
4. Show the classification accuracy per iteration.
5. Show the average classification accuracy after the 10-fold CV is completed.
6. Repeat 1-5 with "corrupted_iris_dataset.dat". This data set is corrupted by Gaussian noise. You should see some reduction in the classification accuracy

Expected output:

“iris_dataset.dat”

```
Classification accuracy
ans =
    0.9333
    1.0000
    1.0000
    1.0000
    1.0000
    1.0000
    1.0000
    0.9333
    0.9333
    1.0000
    0.9333
Average accuracy = 0.9733
```

“corrupted_iris_dataset.dat”

```
Classification accuracy
ans =
    0.8667
    0.8667
    0.9333
    0.7333
    0.8000
    0.8667
    0.8667
    0.7333
    0.7333
    0.8000
Average accuracy = 0.8200
```