

Lab 3

Use Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) over the provided dataset. You will use Python, Numpy, Sci-kit Learn, and Matplotlib for this assignment. Perform classifications on the Iris dataset which can be downloaded at <http://www.cse.scu.edu/~yfang/coen140/iris.data>

The dataset contains 3 classes of 50 instances each, where each class refers to a type of iris plant.

Attribute Information:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
 - Iris Setosa
 - Iris Versicolour
 - Iris Virginica

Exercises:

1. Break the sample into 80% for training, and 20% for test datasets. Choose the first 80% instances from each class for training and the remaining 20% for testing.
 - a. Hint: make sure your initial representation of the data set (of type List[List[]]) passes the [provided test dataset function](#). This is not required, since you can substitute the exact types (np.float64 instead of float, int instead of str), but it is a step in the right direction.
2. Build a LDA classifier based on the training data. [Use the appropriate classifier built into sci-kit learn](#). Report the training and test errors.
 - a. Make a function that returns your trained classifier. Train solely over the training data.

- b. Note that, when passing a numpy array of *samples* into a given classifier's *predict()* function, you may run into an error regarding casting values to *np.float64*. If you run into this, use *samples = samples.astype(np.float64)*.
3. Build a QDA classifier based on the training data. Use the appropriate classifier built into sci-kit learn. Report the training and test errors.
 - a. Train solely over the training data.
4. Are any of the variables not important in classifying iris type? Explain your answer based on your experiments.