

# Capstone Project Report

Name: Sumit Patel

Course: AI and ML (Batch – AUG 2020)

Duration: 10 months

## Linear Discriminant Analysis

### **Problem Statement:**

In order to implement LDA, first generate a dummy dataset (say IRIS dataset having 4 features) and then use LDA to decrease the number of features to one/two. Now using this modified dataset, try to learn a classifier to test the performance of LDA for dimensionality reduction

### **Prerequisites**

What things you need to install the software and how to install them:

Python 3.6 This setup requires that your machine has latest version of python. The following url <https://www.python.org/downloads/> can be referred to download python. Once you have python downloaded and installed, you will need to setup PATH variables (if you want to run python program directly, detail instructions are below in how to run software section). To do that check this: <https://www.pythoncentral.io/add-python-to-path-python-is-not-recognized-as-an-internal-or-external-command/>. Setting up PATH variable is optional as you can also run program without it and more instructions are given below on this topic. Second and easier option is to download anaconda and use its anaconda prompt to run the commands.

To install anaconda check this url <https://www.anaconda.com/download/> You will also need to download and install below 3 packages after you install either python or anaconda from the steps above Sklearn (scikit-learn) numpy scipy if you have chosen to install python 3.6 then run below commands in command prompt/terminal to install these packages:

```
pip install numpy
```

```
pip install matplotlib
```

```
pip install sklearn
```

If you have chosen to install anaconda then run below commands in anaconda prompt to install these packages:

```
conda install -c anaconda numpy
```

```
conda install -c anaconda matplotlib
```

```
conda install -c anaconda sklearn
```

Dataset used:

```
iris = sklearn.datasets.load_iris()
```

Method used for detection

DecisionTreeClassifier(criterion='entropy') and LinearDiscriminantAnalysis

Importing the libraries and loading dataset.

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn import datasets
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from mpl_toolkits import mplot3d
```

Training Data:

```
# Split into train/test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)
```

```
# Use a Decision Tree classifier to evaluate the methods
tree = DecisionTreeClassifier(criterion='entropy')
```

```
tree.fit(X_train,y_train)
acc = tree.score(X_test,y_test)
print("Without LDA",acc)
```

Without LDA 0.9555555555555556

```
# Fit the method's model
lda2.fit(X_train, y_train)

# Fit a Decision Tree classifier on the embedded training set
tree.fit(lda2.transform(X_train), y_train)

# Compute the Decision Tree accuracy on the embedded test set
acc = tree.score(lda2.transform(X_test), y_test)
X_2 = lda2.transform(X)
plt.figure(figsize=(12,5))
for color, i, target_name in zip(colors, [0, 1, 2], target_names):
    plt.scatter(X_2[y == i, 0], X_2[y == i, 1], alpha=.8, color=color,
                label=target_name)
plt.legend(loc='best')
plt.title("LDA, Decision Tree\nTest accuracy = {:.2f}".format(acc))
plt.show()
```

Test Results

