

# ASSIGNMENT :- NAIVE BAYES & LDA MODEL COMPARISON ¶

In [2]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import model_selection
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.naive_bayes import GaussianNB
```

In [3]:

```
# Load dataset
dataset = pd.read_csv('C:/Users/ddalv/Documents/Courses/ML & Stats/HW/LDA/Wine.csv')
dataset.head()
```

Out[3]:

	Alcohol	Malic_Acid	Ash	Ash_Alcanity	Magnesium	Total_Phenols	Flava
0	14.23	1.71	2.43	15.6	127	2.80	3.06
1	13.20	1.78	2.14	11.2	100	2.65	2.76
2	13.16	2.36	2.67	18.6	101	2.80	3.24
3	14.37	1.95	2.50	16.8	113	3.85	3.49
4	13.24	2.59	2.87	21.0	118	2.80	2.69

In [4]:

```
X = dataset.iloc[:, 0:13].values
Y = dataset.iloc[:, 13].values
```

In [5]:

```
# prepare configuration for cross validation test harness

seed = 10
```

In [6]:

```
# prepare models

models = []
models.append(('LDA', LinearDiscriminantAnalysis()))
models.append(('NB', GaussianNB()))
```

In [7]:

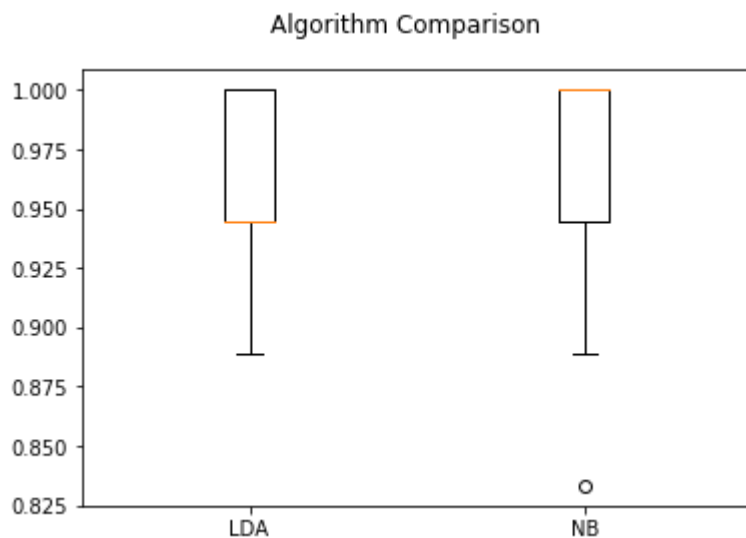
```
# evaluate each model in turn
results = []
names = []
scoring = 'accuracy'
for name, model in models:
    kfold = model_selection.KFold(n_splits=10, random_state=seed)
    cv_results = model_selection.cross_val_score(model, X, Y, cv=kfold, scoring=sco
ring)
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    print(msg)
```

LDA: 0.961111 (0.035573)

NB: 0.961111 (0.055833)

In [8]:

```
# boxplot algorithm comparison
fig = plt.figure()
fig.suptitle('Algorithm Comparison')
ax = fig.add_subplot(111)
plt.boxplot(results)
ax.set_xticklabels(names)
plt.show()
```



**Both models are performing equally well on the dataset.**