In [1]:
```
"""
ECGR 5105 - Intro to Machine Learning
Homework 2 - Part 2
Phillip Harmon
"""
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

In [2]:
```
#Import Dataset
csvData = pd.read_csv('diabetes.csv')
csvData
```

Out[2]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunctio
0	6	148	72	35	0	33.6	0.62
1	1	85	66	29	0	26.6	0.35
2	8	183	64	0	0	23.3	0.67
3	1	89	66	23	94	28.1	0.16
4	0	137	40	35	168	43.1	2.28
...	...	...	...	...	...	...	
763	10	101	76	48	180	32.9	0.17
764	2	122	70	27	0	36.8	0.34
765	5	121	72	23	112	26.2	0.24
766	1	126	60	0	0	30.1	0.34
767	1	93	70	31	0	30.4	0.31

768 rows × 9 columns

In [3]:
```
#Sort Dataset
x = csvData.iloc[:,0:-1].values
y = csvData.iloc[:,-1].values
```

In [18]:
```
#Clean the Dataset
from sklearn.preprocessing import MinMaxScaler, StandardScaler
scaler = StandardScaler()
 # scaler = MinMaxScaler() #StandardScaler gave better results here
x = scaler.fit_transform(x)
```

In [19]:
```
#Perform the Training with K=5
from sklearn.model_selection import KFold
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import cross_val_score
kcup = KFold(n_splits=5, random_state=1337, shuffle=True)
model = LogisticRegression(random_state=1337)
results = cross_val_score(model,x,y,cv=kcup)
print("K=5 | Accuracy: {:.3f}% ({:.3f}%)".format(results.mean()*100, results.s
```

```
K=5 | Accuracy: 77.734% (2.167%)
```

In [17]:
```
#Perform the Training with K=10
kcup = KFold(n_splits=10, random_state=1337, shuffle=True)
model = LogisticRegression(random_state=1337)
results = cross_val_score(model,x,y,cv=kcup)
print("K=10 | Accuracy: {:.3f}% ({:.3f}%)".format(results.mean()*100, results.
```

```
K=10 | Accuracy: 76.811% (3.108%)
```

In [ ]:
```

```