

Assignment on Machine Learning I_Laxman

Split the Heart dataset into a training set (80%) and test set (20%)

Use the training set to fit classification models in order to predict whether a person is likely to suffer from Arterial Heart Disease (AHD). Consider the following classification techniques:

1. Logistic Regression with L2 regularization

Confusion Matrix:

```
[[29  1]
 [ 4 26]]
```

	precision	recall	f1-score	support
0	0.88	0.97	0.92	30
1	0.96	0.87	0.91	30
accuracy			0.92	60
macro avg	0.92	0.92	0.92	60
weighted avg	0.92	0.92	0.92	60

The accuracy score achieved using Logistic Regression is: 91.67 %

2. Linear Discriminant Analysis

Threshold = 0.9

Confusion Matrix:

```
[29,  1],
 [ 2, 28]])
```

	precision	recall	f1-score	support
0	0.94	0.97	0.95	30
1	0.97	0.93	0.95	30
accuracy			0.95	60
macro avg	0.95	0.95	0.95	60
weighted avg	0.95	0.95	0.95	60

The accuracy score achieved using LDA is: 95.0 %

Threshold = 0.9

Confusion Matrix:

```
([[29,  1],
 [ 2, 28]])
```

	precision	recall	f1-score	support
0	0.94	0.97	0.95	30
1	0.97	0.93	0.95	30
accuracy			0.95	60
macro avg	0.95	0.95	0.95	60
weighted avg	0.95	0.95	0.95	60

The accuracy score achieved using LDA is: 95.0 %

3. K-Nearest neighbors

K=3

Confusion Matrix:

```
[[17 13]
 [ 7 23]]
```

	precision	recall	f1-score	support
0	0.71	0.57	0.63	30
1	0.64	0.77	0.70	30
accuracy			0.67	60
macro avg	0.67	0.67	0.66	60
weighted avg		0.67	0.67	0.66 60

The accuracy score achieved using KNN is: 66.67 %

K=5

Confusion Matrix:

```
[[15 15]
 [ 9 21]]
```

	precision	recall	f1-score	support
0	0.62	0.50	0.56	30
1	0.58	0.70	0.64	30
accuracy			0.60	60
macro avg	0.60	0.60	0.60	60
weighted avg	0.60	0.60	0.60	60

The accuracy score achieved using KNN is: 60.0 %

4. Naïve Bayes

```
Confusion Matrix:
[[25  5]
 [ 6 24]]
      precision    recall  f1-score   support

     0       0.81      0.83      0.82         30
     1       0.83      0.80      0.81         30

 accuracy          0.82          0.82          0.82         60
 macro avg       0.82      0.82      0.82         60
weighted avg       0.82      0.82      0.82         60

The accuracy score achieved using Naive Bayes is: 81.67 %
```

Use **10-fold cross validation over the training set (80%)** for training the models. **Generate the confusion matrix for the test data.** Also calculate the **precision, recall, F1 and accuracy scores.** Train your model for **different threshold values for the LDA and different K values for the KNN** and for the test data use the model that produced the best accuracy.

Write your codes in python and submit the .pynb files **along with a report** on the above mentioned results on the test data.