CO 544 Machine Learning and Data Mining

Lab 01

E/15/366

8. Note down the results in the below table.

	Correctly classified instances	Incorrectly classified instances
Training Set	143 (92.2581%)	12 (7.7419%)
Cross validation(10 folds	130 (83.871%)	25 (16.129%)

9. Interpret the results from the 'confusion matrix' in the classifier output.

Cross validation

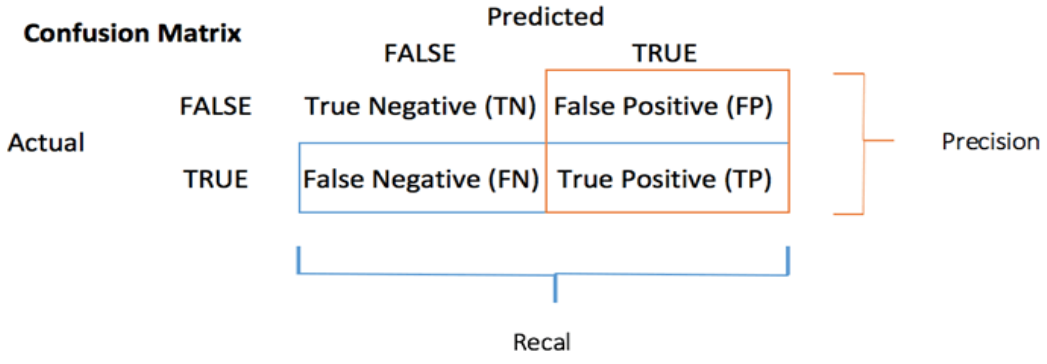
Confusion Matrix

a b -- classified as
14 18 | a = DIE
7 116 | b = LIVE

Use of training data

Confusion Matrix

a b -- classified as
22 10 | a = DIE
2 121 | b = LIVE



Confusion Matrix is a useful machine learning method which allows you to measure Recall, Precision, Accuracy, and AUC-ROC curve.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

	Accuracy	Precision	Recall
Training Set	0.9225	0.9236	0.9837
Cross validation	0.8387	0.8656	0.9430

When using cross validation the accuracy decreased because cross validation divide the data set into n number of parts, run the algorithm and take the average of it, so the accuracy shown here is more realistic than the usual training set.

To get the value of precision we divide the total number of correctly classified positive examples by the total number of predicted positive examples. High Precision indicates an example labelled as positive is indeed positive.

Recall can be defined as the ratio of the total number of correctly classified positive examples divide to the total number of positive examples. High Recall indicates the class is correctly recognized.

10. Change the parameters as below and compare the results with the results of the above model with default values. Confidence factor: 0.5 and Min number of folds:2.

Training set

Correctly Classified Instances	149	96.129 %
Incorrectly Classified Instances	6	3.871 %

Confusion Matrix

a	b	← classified as
28	4	a = DIE
2	121	b = LIVE

Accuracy = 96.129%

The accuracy is increased when compared to Training set default values.

Tree size also increased when the values are changed, in default values tree size is 21 and number of leaves are 11 after changed the size of tree increased to 31 and number of leaves increased to 16.

Cross validation

Correctly Classified Instances	126	81.2903 %
Incorrectly Classified Instances	29	18.7097 %

Confusion Matrix

a	b	← classified as
18	14	a = DIE
15	108	b = LIVE

Accuracy = 81.2903%

The accuracy is decreased when compared to Cross validation default values.

Tree size also increased when the values are changed, in default values tree size is 21 and number of leaves are 11 after changed the size of tree increased to 31 and number of leaves increased to 16.