

# CO544 Machine Learning and Data Mining - Lab01

E/15/140  
Jaliyagoda A.J.N.M.

8. Note down the results in the below table.

	Correctly classified instances	Incorrectly classified instances
Training Set	143	12
Cross-validation(10 folds)	130	25

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

Following is the standard format of a confusion matrix.

Actual \ Predicted	Class 1	Class 2
Class 1	TP: True Positive	FN: False Negative
Class 2	FP: False Positive	TN: True Negative

Training Set:

Actual \ Predicted	a	b
a	22	10
b	2	121

a=DIE, b=LIVE

Cross-validation(10 folds)

Actual \ Predicted	a	b
a	14	18
b	7	116

a=DIE, b=LIVE

A confusion matrix is a table that is often used to describe the performance of a classification model (or “classifier”) on a set of test data for which the true values are known. It allows the visualization of the performance of an algorithm. It allows easy identification of confusion between classes e.g. one class is commonly mislabeled as the other. Most performance measures are computed from the confusion matrix.

There are few measures, which can be used to interpret the result of the confusion matrix

1. Classification Rate/Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Recall

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

3. Precision

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

Here are the calculated values for the above 3 measures under Training set and Cross-validation with 10 folds. (DIE is considered as a positive result)

	Training Set	Cross-validation(10 folds)
Accuracy	13/15 = 0.867	26/31 = 0.839
Recall	11/16 = 0.688	7/16 = 0.438
Precision	11/12 = 0.917	2/3 = 0.667

**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**

Confidence Factor: 0.2

Min Number: 2

Correctly Classified Instances      149              96.129 %

Incorrectly Classified Instances      6              3.871 %

Actual \ Predicted	a	b
a	28	4
b	2	121

	Training Set (default values)	CF=0.5, minFold=2
Accuracy	13/15 = 0.867	149/155 = 0.961
Recall	11/16 = 0.688	7/8 = 0.857
Precision	11/12 = 0.917	14/15 = 0.933

## REFERENCES

1. <https://machinelearningmastery.com/k-fold-cross-validation/>
2. <https://www.geeksforgeeks.org/confusion-matrix-machine-learning/>