

Data Analytics with Python

Week 9 Assignment Solution

Q1. The following confusion matrix was obtained from a classifier. Confusion Matrix:

	True Class			
		Apple	Orange	Mango
	Apple	7	8	9
	Orange	1	2	3
	Mango	3	2	1

What is the accuracy of the classifier?

$$\text{Accuracy} = \frac{\text{Number of Correct Prediction (Digonal elements)}}{\text{Total number of Predictions}}$$

$$\text{Accuracy} = \frac{7 + 2 + 1}{7 + 8 + 9 + 1 + 2 + 3 + 3 + 2 + 1}$$

$$\text{Accuracy} = \frac{10}{36}$$

$$\text{Accuracy} = 0.27778 = 27\%$$

Q2. For the given confusion matrix, what is the number of False Positives for the Apple class?

	True Class			
		Apple	Orange	Mango
Predicted Class	Apple	7	8	9
	Orange	1	2	3
	Mango	3	2	1

False Positive means the label does not belong to the class, but classifier predicted as positive.

FP for Apple = (predicted as Apple but actually Orange) + (predicted as Apple but actually Mango)

$$= 8 + 9$$

$$= 17$$

Answer: False Positive for Apple = 17

Q3. For the given confusion matrix, what is the number of True Negative for the Apple class?

	True Class			
		Apple	Orange	Mango
	Apple	7	8	9
	Orange	1	2	3
	Mango	3	2	1

True Negative means that the label does not belong to the class, and it is correctly predicted.

$$\text{TN for Apples} = 2 + 3 + 2 + 1 = 8$$

Answer: TN for Apple = 8

Q4. For the given confusion matrix, what is the number of False Negative for the Apple class?

	True Class			
		Apple	Orange	Mango
	Apple	7	8	9
	Orange	1	2	3
	Mango	3	2	1

False Negative means that the label belong to the class but it predicted negative.

FN for Apples = Belong to Apple but predicted as Orange and Mango

$$= 1 + 3$$

$$= 4$$

Answer: FN for Apple = 4

Q5. For the given confusion matrix, what is the F1-score of the classifier for the Apple class?

	True Class			
		Apple	Orange	Mango
Predicted Class	Apple	7	8	9
	Orange	1	2	3
	Mango	3	2	1

For Apple Class

$$TP = 7$$

$$TN = 2 + 3 + 2 + 1 = 8$$

$$FP = 8 + 9 = 17$$

$$FN = 4$$

$$Precision = \frac{TP}{TP + FP} = \frac{7}{7 + 17} = 0.29 \quad \& \quad Recall = \frac{TP}{TP + FN} = \frac{7}{7 + 4} = 0.64$$

$$F1 \text{ Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0.29 \times 0.64}{0.29 + 0.64} = 0.40$$

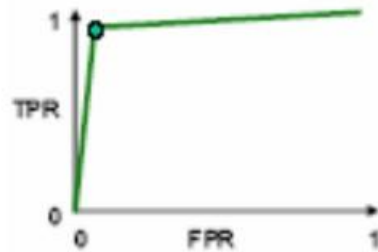
Answer : F1 Score = 0.4

Q6. In ROC analysis, a classifier is called 'good' if it has _____

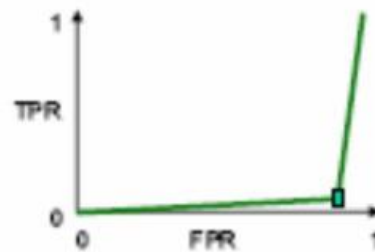
ROC analysis

- ROC = receiver operator/operating characteristic/curve

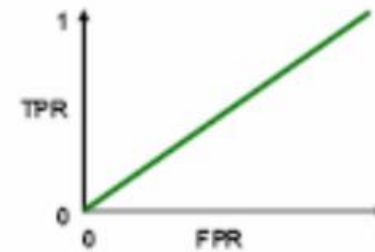
- ROC space: good and bad classifiers.



- Good classifier.
 - High TPR.
 - Low FPR.



- Bad classifier.
 - Low TPR.
 - High FPR.



- Bad classifier (real picture).

Answer: High TPR and Low FPR

Q7. State True or False: Standardization of features is not required before training a Logistic regression model

- A. True
- B. False

Answer: True

Standardization is not required for logistic regression. The main goal of standardizing features is to help convergence of the technique used for optimization.

Q8. For the given confusion matrix, determine the sensitivity for the model.

Predicted	Actual		
		Disease	<u>Non disease</u>
	Positive	10	40
	Negative	5	45

Here

$$TP = 10$$

$$TN = 45$$

$$FP = 40$$

$$FN = 5$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{10}{10 + 5} = 0.6667 \approx 67\%$$

Answer: Sensitivity = 67%

Q9. For the given confusion matrix, determine the specificity for the model.

Predicted	Actual		
		Disease	<u>Non disease</u>
	Positive	10	40
	Negative	5	45

Here

$$TP = 10$$

$$TN = 45$$

$$FP = 40$$

$$FN = 5$$

$$Specificity = \frac{TN}{TN + FP} = \frac{45}{45 + 40} = 0.5294 \approx 53\%$$

Answer: Sensitivity = 53%

10. According to the ROC Curve and AUC below, choose the correct alternative for the effectiveness of classifiers A and B.

The AUC for A ROC curve is higher than that for the B ROC curve.

Therefore, We can say that A did a better job of classifying the positive class in the dataset

Answer: $A > B$

