Measuring Classifier Performance

8 performance metrics to evaluate a classifier.

1) Accuracy
2) Confusion Matrix / Contingency table
3) AUC / ROC
4) Precision
5) Recall
6) F1 Score

1) Accuracy:

Accuracy of a model is simply the number of correct predictions divided by the total number of predictions. An accuracy score will be between the range 0 & 1. The value of 1 indicate a perfect model.

$$Accuracy = Correct\ Predictions\ /\ Total\ Predictions$$

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ positive + False\ Negative}$$

2) Confusion Matrix

It is a tool to observe in which way the model is wrong / right. It is a matrix that compares the number of predictions for each class that are correct & those that are incorrect.

True Positives → The number of positive observations the model correctly predicted as positive.

True positive → You predicted positive, & it's true

True Negative → You predicted Negative & it's true. The number of negative observations. the model correctly predicted as negative.

False positive → also called as (Type 1 Error). You predicted positive & it's false. In other words, the number of negative observations the model incorrectly predicted as positive.

False Negative → also called as (Type 2 Error) You predicted negative & it is false. In other words, the number of positive observations, the model incorrectly predicted as negative.

True class 0	3383 Negative sample	46. Negative Sample as Positive
1	89 Positive Sample as negative	982 Positive Sample
	0	1

Predicted class

In the above figure, the model correctly predicted 3383 negative samples but incorrectly predicted 46 samples as positive.

The model correctly predicted 982 positive observations but incorrectly predicted 89 as negative.

True positive rate [TPR] is also called as sensitivity (or) recall.

$$\text{TPR / Sensitivity / Recall} = \frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

The proportion of actual positive cases which are correctly identified.

True Negative rate [TNR] is also called as specificity

$$\text{TNR / Specificity} = \frac{\text{True Negative}}{\text{True Negative + False Positive}}$$

The proportion of actual negative cases which are correctly identified.

Precision → Precision is the ratio between the true positives & all the positives.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

F_1 Score → F_1 Score is the harmonic mean between precision & recall.

$$F_1 \text{ Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision + Recall}}$$

1) Problem: 1

Consider the following 3 class confusion matrix. calculate precision & recall per class. Also calculate weighted average precision & recall for classifier.

	Predicted		
	Class 1	Class 2	Class 3
	15	2	3
Actual	7	15	8
	2	3	45

Solut

	Predicted			
Actual	15	2	3	**20** $(15+2+3)=20$
	7	15	8	**30** $(7+15+8)=30$
	2	3	45	**50** $(2+3+45)=50$
	24	**20**	**56**	**100**
	$(15+7+2)$	$(2+15+3)$	$(3+8+45)$	$(20+30+50)$

$$\text{Accuracy} = \frac{15 + 15 + 45}{100} = \frac{75}{100} = 0.75$$

i) Precision & recall per class.

Class 1 $= \frac{15}{24} = 0.63$ & $\frac{15}{20} = 0.75$

Class 2 $= \frac{15}{20} = 0.75$ & $\frac{15}{30} = 0.50$

Class 3 $= \frac{45}{56} = 0.8$ & $\frac{45}{50} = 0.9$

2) **Problem 2:-** Consider following confusion matrix & calculate following i) Sensitivity of classifier, ii) Specificity of classifier.

confusion matrix		Predicted		Total
Actual	+	8	10	18
	–	4	8	12
Total		12	18	30

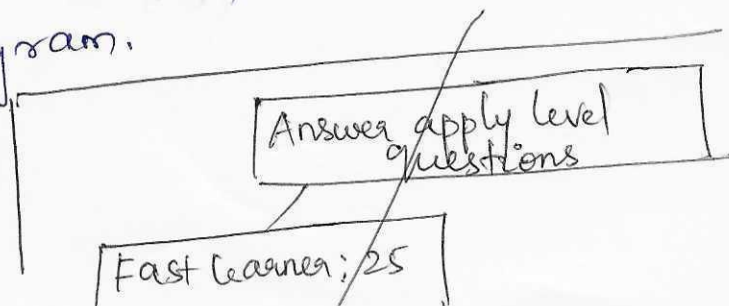**Soln:-**
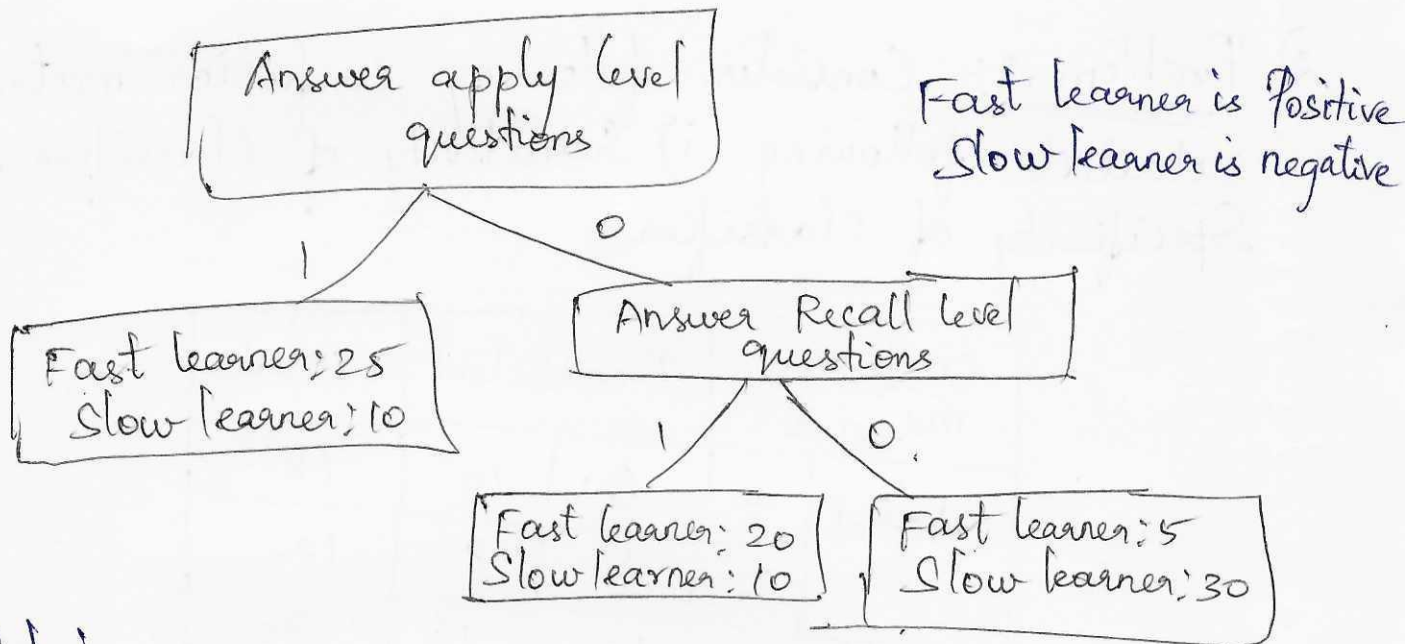
True positive is 8, True negative is 8, False positive is 10 & False negative is 4.

i) Sensitivity $= \dfrac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$

$\qquad = \dfrac{8}{8+4} = 8/12 = 0.666$

ii) Specificity $= \dfrac{\text{True Negative}}{\text{True Negative} + \text{False positive}}$

$\qquad = \dfrac{8}{8+10} = \dfrac{8}{18} = 0.444$

3) **Problem 3:-** Find contingency table, recall, Precision & Negative recall, False positive rate for the below diagram.

Answer apply level questions

Fast Learner:/25

Answer apply level questions

Fast learner is Positive
Slow learner is negative

1 | 0

Fast learner: 25
Slow learner: 10

Answer Recall level questions

1 | 0

Fast learner: 20
Slow learner: 10

Fast learner: 5
Slow learner: 30

Soln:

Contingency table.

	Predicted			Total
Fast	25	20	5	50
Slow	10	10	30	50
Total	35	30	35	100

$$\text{Precision} = \frac{\text{True Positive}}{\text{Actual Results}} = \frac{25 \ (\text{Fast learner})}{35 \ (\text{Total learners})} = 0.714.$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{Predicted results}} = \frac{25}{30} = 0.833$$

$$\text{False positive rate} = \frac{\text{False positive}}{FP + TN} = \frac{10}{10 + 30} = 0.25.$$