

* Evaluation of Logistic Regression model

- Error metrics frequently used in industry are as follows,

A.1. Confusion matrix (Error matrix)

- Matrix with count of various types of prediction
- **True Positive (TP)** model predicted yes and correct answer was yes.
- **True Negative (TN)** model predicted no and correct answer was no.
- **False Positive (FP)** model predicted yes and correct answer was no.
- **False Negative (FN)** model predicted no and correct answer was yes.

Note: Was model right or wrong + what model predicted

- This is how TP, TN, FP, FN works.
- Based on confusion matrix we have certain error calculation metrics.

A.1. Accuracy (A)

- Tells overall how classifier is correct

$$A = \frac{TP + TN}{\text{Total}}$$

A.2. Recall or Sensitivity (R)

- True Positive rate
- it says about how much time yes is predicted correctly.
- fraction of patterns correctly predicted.

$$R = \frac{TP}{\text{actual yes}}$$

$$\Rightarrow R = \frac{TP}{TP + FN}$$

A.3. Precision (P)

- if it is predicted yes then, how often it is correct?
- number positive patterns predicted correctly, by total number of patterns in positive class.

$$P = TP / \text{predicted yes}$$

$$\Rightarrow P = \frac{TP}{TP + FP}$$

A.4. Misclassification Rate (MR)

- Error rate

$$MR = \frac{FP + FN}{\text{Total}}$$

A.5. Specificity (S)

- True Negative rate
- it tells about how much time no is predicted correctly and actual value is also no.

$$S = TN / \text{actual no}$$

$$\Rightarrow S = \frac{TN}{TN + FP}$$

$$\text{also } S = 1 - R$$

A.6. F-β score

- weighted harmonic mean of precision and recall reaching its optimal value at 1 and worst value at 0.
- β determines weight of recall.
- In general,

$$F-\beta \text{ score} = \frac{(1 + \beta^2) [P \times R]}{\beta^2 \times P + R}$$

@ If FP and FN both are important,
 $\beta = 1$

$$F1 \text{ score} = \frac{2 \times PR}{P + R}$$

⑥ If FP more important than FN,
 $\beta = 0.5$

$$F_{0.5} \text{ score} = \frac{1.25 \times PR}{0.25P + R}$$

⑦ If FN more important than FP,
 $\beta = 2$

$$F2 \text{ score} = \frac{5PR}{4P + R}$$