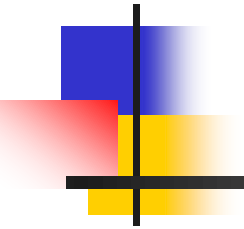
Model Evaluation Parameters for Regression and Classification



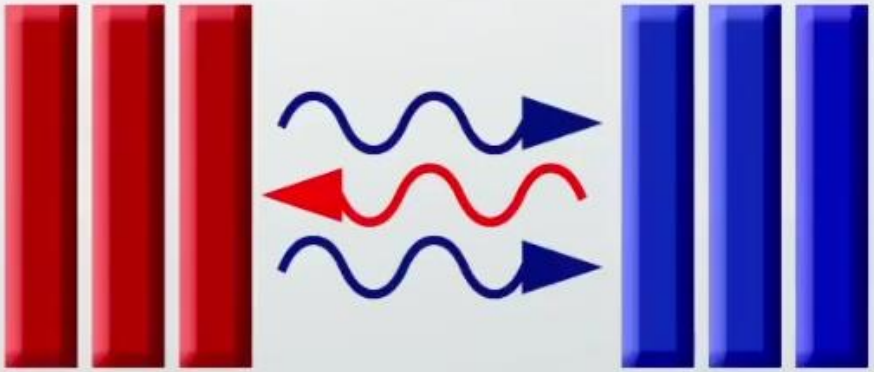
Dr. Prashant Singh Rana
psrana@gmail.com

Evaluation Parameters for Regression

- Parameters on which we compare different models and system.
  - Performance parameters:
    - Correlation
    - R^2
    - Accuracy
    - Prediction Time
    - Error (RMSE, MAE, MSA, etc)

1. Correlation: Pearson

the relationship between two sets of variables used to describe or predict information



1. Correlation: Pearson

Formula

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

1. Correlation: Pearson

- The closer the number is to positive one, the stronger the positive correlation
- The closer the number is to negative one, the stronger the negative correlation
- The closer the number is to zero, the weaker the correlation

-1

- Zero means there is no correlation between the variables

1

2. Coefficient of determinant

- The coefficient of determination (R2) summarizes the **explanatory power of the regression model**.

- R2 describes the **proportion of variance of the dependent variable explained by the regression model**.

- If the regression model is perfect then R2 is 1 and if the regression model is a total failure then R2 is zero i.e. no variance is explained by regression.

- It is defined as follows: $R^2 = r * r$

3. Error

Mean squared error	$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2$
Root mean squared error	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$
Mean absolute error	$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} \lvert e_t \rvert$

Where, e is the error between x and y

4. Accuracy

- The accuracy is calculated as percentage deviation of predicted target with actual target (with or without **acceptable error**).

- Example:
  - Dose given by doctor (100 mg +-10) → **OK**
  - Dose given by doctor (100 mg +-50) → **Not OK**

4. Accuracy

$$Accuracy = \frac{100}{n} \sum_{i=1}^{n} q_i$$

$$q_i = \begin{cases} 1 & if abs(p_i - a_i) \leq err \\ 0 & otherwise \end{cases}$$

4. Accuracy: Find

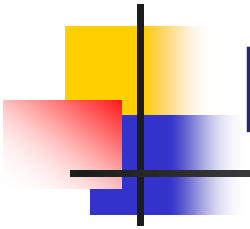
Accuracy with +-5 error

- = IF (ABS (A4 - B4) <= 5, 1, 0)
- = AVERAGE (C4 : C16) * 100

4. Accuracy: Find

		Accuracy			
		With +- 5 err	With +- 10 err	With +- 20 err	With +- 30 err
Actual	Predicted	23.08	38.46	76.92	92.31
24	21	1	1	1	1
50	48	1	1	1	1
42	54	0	0	1	1
51	32	0	0	1	1
11	32	0	0	0	1
24	11	0	0	1	1
32	38	0	1	1	1
27	56	0	0	0	1
59	54	1	1	1	1
56	49	0	1	1	1
45	29	0	0	1	1
16	59	0	0	0	0
56	37	0	0	1	1

Evaluation Parameters for Binary Classification

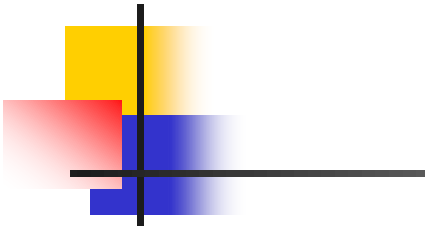


Binary Classification TRUE/FALSE

Actual	Predicted
1	0
1	1
0	1
1	0
0	1
0	1
0	1
0	0
1	0
0	1

Confusion Matrix / Error Matrix

n= 10	**Predicted: YES**	**Predicted: NO**	
Actual: YES	TP = 1	FN = 3	4
Actual: NO	FP = 5	TN = 1	6
	6	4	



$$\text{Sensitivity or recall} = \frac{TP}{(TP + FN)}$$

$$\text{Specificity} = \frac{TN}{(TN + FP)}$$

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

$$\text{Positive Predictive Value or Precision} = \frac{TP}{(TP + FP)}$$

$$\text{Negative Predictive Value} = \frac{TN}{(TN + FN)}$$

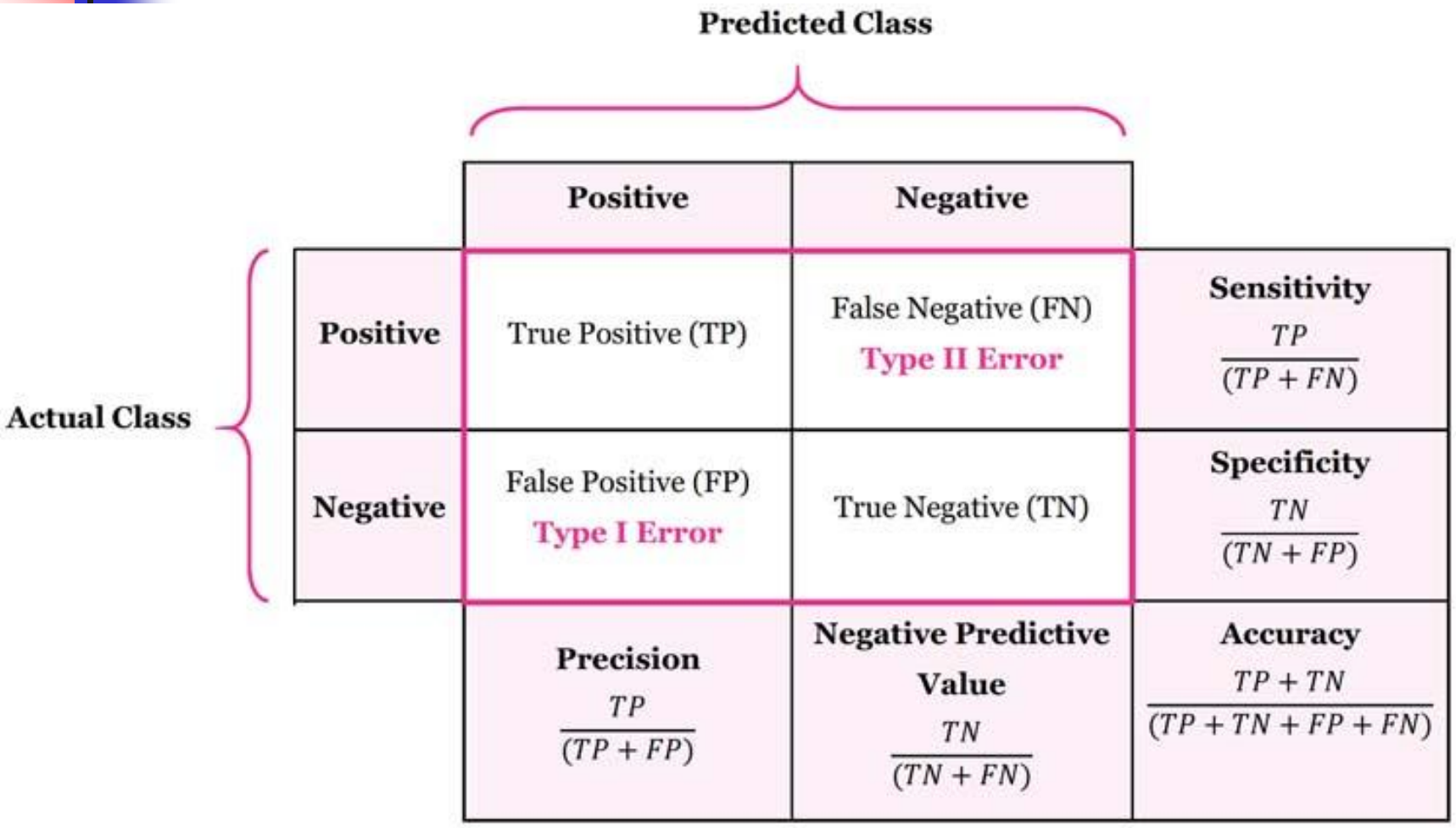
$$\text{F-measure} = 2 * \frac{(precision * recall)}{(precision + recall)}$$

Sensitivity / Specificity

- **Sensitivity** presents **true positive rate.**
  - **1 → 1**
  - **Cancer → Cancer**

- **Specificity** presents **true negative rate**.
  - **0 → 0**
  - **Non-Cancer → Non-Cancer**

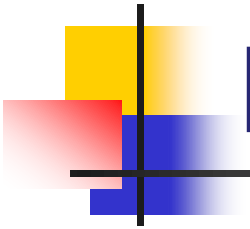
Confusion Matrix / Error Matrix



Question:
Calculate all parameters for

Actual	Predicted
1	0
1	1
0	1
1	0
0	1
0	1
0	1
0	0
1	0
0	1

Evaluation Parameters for Multi-Class Classification



Model Evaluation Parameters

- Error Matrix

Actual/ Predicted	0	1	2	3	4	5
0	1305	1	5	15	15	5
1	0	114	28	0	0	0
2	4	31	551	136	17	2
3	8	2	125	786	150	13
4	22	3	39	206	713	104
5	14	2	17	46	121	314

Model Evaluation Parameters

- For Accuracy

$$Accuracy = \frac{100}{n} \sum_{i=1}^{n} q_i$$

$$q_i = \begin{cases} 1 & p_i = a_i \\ 0 & otherwise \end{cases}$$

Model Evaluation Parameters

- Calculate error matrix and Accuracy

Actual	Predicted
2	0
1	2
2	4
0	4
3	4
0	4
0	0
0	2
3	4
0	0
4	4
4	4

Class wise error

		True/Actual		
		Cat ( )	Fish ( )	Hen ( )
Predicted	Cat ( )	4	6	3
	Fish ( )	1	2	0
	Hen ( )	1	2	6

Number of correctly classified: 4 + 2 + 6 = 12

Total number = 25

- **Overall Accuracy = 12/25 = 48.0%**

- **Error (Cat): (1 + 1) / (4 + 1 + 1) = 2 / 6 = 0.33**
- **Error (Fish): (6 + 2) / (6 + 2 +2) = 8 / 10 = 0.80**
- **Error (Hen): (3 + 0) / (3 + 0 + 6) = 3 / 9 = 0.33**

Thank You