

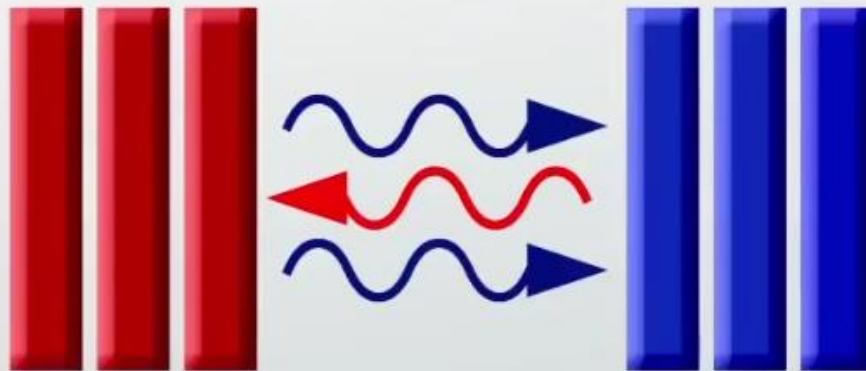


Evaluation Parameters for Regression

- 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

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

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



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 (R^2) summarizes the **explanatory power of the regression model**.
- R^2 describes the **proportion of variance of the dependent variable explained by the regression model**.
- If the regression model is perfect then R^2 is 1 and if the regression model is a total failure then R^2 is zero i.e. no variance is explained by regression.
- It is defined as follows: $R^2 = r * r$



3. Error

Mean squared error

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n e_t^2$$

Root mean squared error

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2}$$

Mean absolute error

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |e_t|$$

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) \rightarrow **OK**
 - Dose given by doctor (100 mg ± 50) \rightarrow **Not OK**



4. Accuracy

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

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



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	


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

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

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

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

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

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



Sensitivity / Specificity

- **Sensitivity** presents **true positive rate**.
 - $1 \rightarrow 1$
 - Cancer \rightarrow Cancer
- **Specificity** presents **true negative rate**.
 - $0 \rightarrow 0$
 - Non-Cancer \rightarrow Non-Cancer

Confusion Matrix / Error Matrix

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$



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):** $(3 + 0) / (3 + 0 + 6) = 3 / 9 = 0.33$