Model Evaluation Parameters

Dr. SINGARA SINGH KASANA
ASSOCIATE PROFESSOR, CSED
TIET, PATIALA

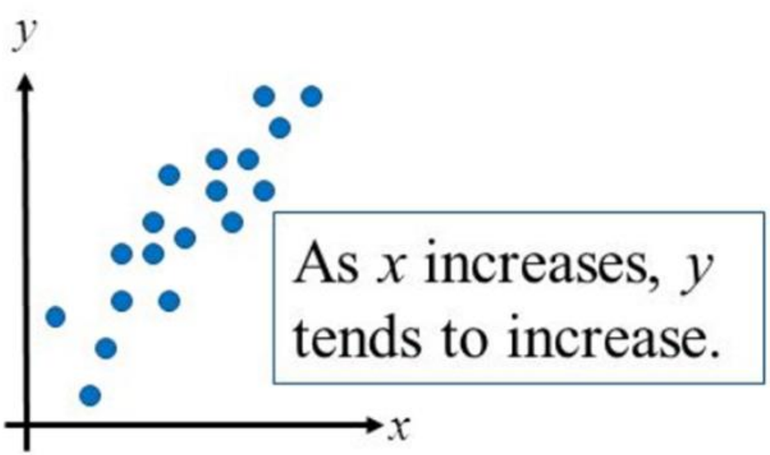
- The idea of building machine learning models works on a constructive feedback principle.
- You build a model, get feedback from metrics, make improvements and continue until you achieve a desirable accuracy.
- Evaluation metrics explain the performance of a model.
- The important aspect of evaluation metrics is their capability to discriminate among model results.

Evaluation Parameters for Regression Models:

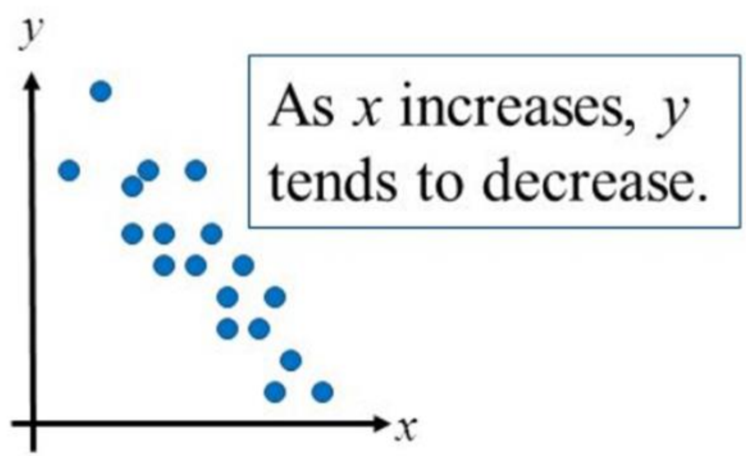
- Correlation
- R^2
- MSE
- RMSE
- MAE
- Accuracy

1. Correlation

Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together.
Correlation varies between -1 to 1.



Positive Linear Correlation



Negative Linear Correlation

- Pearson Correlation

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

Actual Values	**Predicted Values**
5.1	3.5
4.9	3.0
4.7	3.2
4.6	3.1
5.0	3.6

Correlation between the actual and predicted values is ??

2. R^2 : Coefficient of determination

$$R^2 = r * r$$

- 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.

3. MSE

- Mean Square Error

$$\text{MSE} = \frac{\sum_{i=1}^{n} (p_i - a_i)^2}{n}$$

Where **p** is predicted value, **a** is actual value and **n** is total number of observations.

4. RMSE

- Root Mean Square Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (p_i - a_i)^2}{n}}$$

Where **p** is predicted value, **a** is actual value and **n** is total number of observations.

5. Accuracy

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

Accuracy

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

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

Find the accuracy by taking different values of Error

Actual Output	Predicted Output
20	21
50	54
42	43
24	21
32	29
27	23
45	38
16	46
56	48
59	55

Model Evaluation Parameters

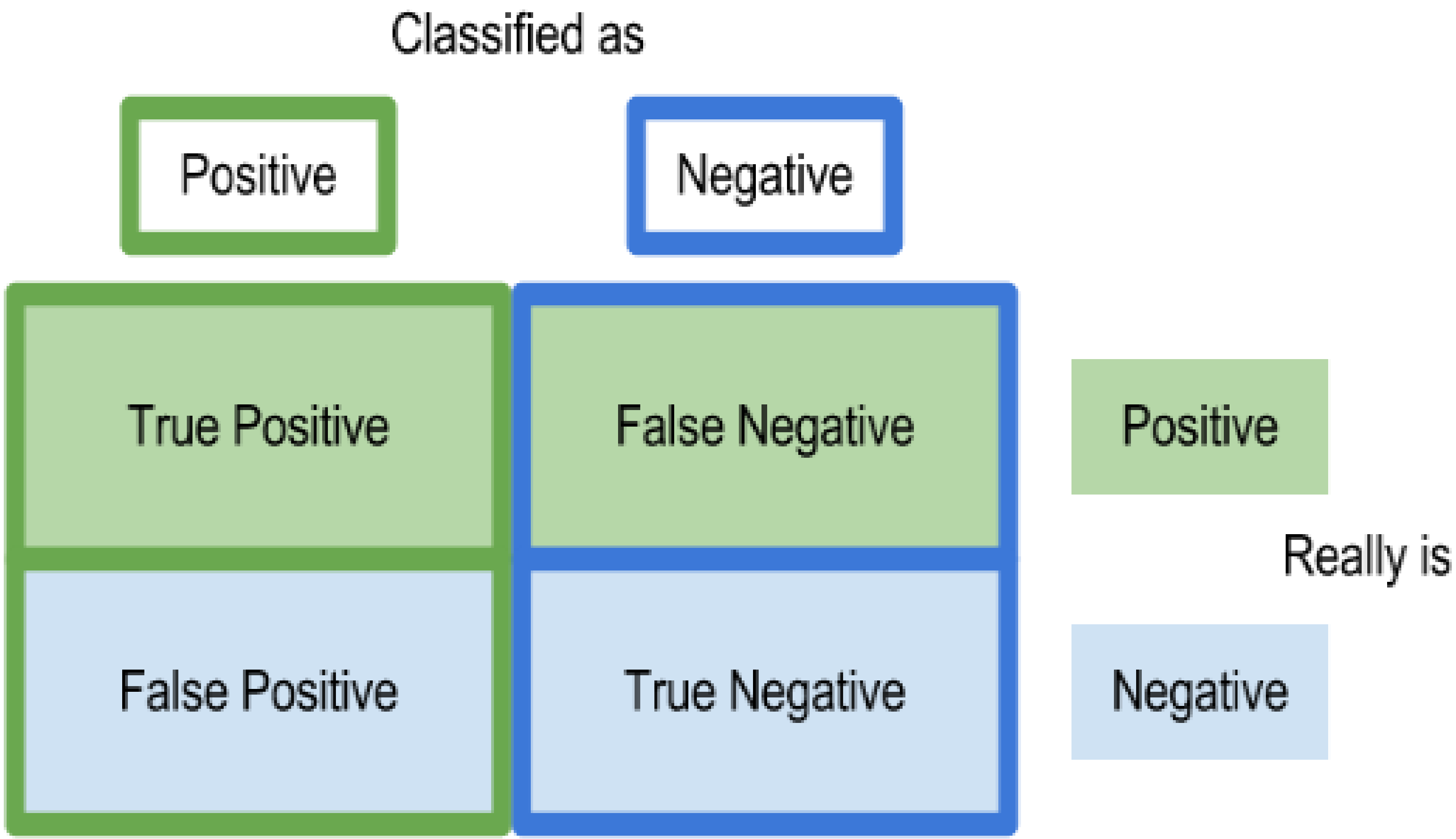
- For Classification

Confusion Matrix

- True Positive(TP)
- True Negative(TN)
- False Positive(FP) – Type I Error
- False Negative(FN) - Type II Error

- Sometimes it is not possible to minimize both the error. So it depends upon the application that which error is more important to minimize.
- Consider an application of classifying a patient as cancer or non-cancer patient. Diagnosing a cancer patent as non cancer(FN) is more important than non-cancer patent as cancer (FP). So FN must be minimized in this case.
- Consider email classification app which classifies mail as spam or ham. Classifying a ham mail as spam (FP) is more important than spam as ham(FN). So FP must be minimized.

Confusion Matrix / Error Matrix



Confusion Matrix

n=165	**Predicted: NO**	**Predicted: YES**	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

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

$$\boxed{\begin{gathered} Accuracy = \frac{100}{n}\sum_{i=1}^{n} q_i \\ q_i = \begin{cases} 1 & p_i = a_i \\ 0 & otherwise \end{cases} \end{gathered}}$$

Classification accuracy is used when dataset is balanced.

Precision tells us how many, out of all instances that were predicted to belong to class C1, actually belonged to class C1 The precision for class C1 is calculated as:

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

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

Also known as micro precision

Macro & Weighted Precision

$$Precision_{positive} = \frac{Correct\ positive\ prediciton}{Total\ positive\ prediction} = \frac{TP}{TP + FP}$$

$$Precision_{negative} = \frac{Correct\ negative\ prediciton}{Total\ negative\ prediction} = \frac{TN}{TN + FN}$$

$$Macro\ Precision = \frac{Precision_{positive} + Precision_{negative}}{2}$$

$$Weighted\ Precision = \frac{n_1 \times Precision_{positive} + n_2 \times Precision_{negative}}{n_1 + n_2}$$

n_1 is number of positive class and n_2 is number of support examples for negative class.

Sensitivity or Recall (True Positive Rate)

It corresponds to the **proportion** of positive data points that are correctly considered as positive, with respect to all positive data points.

$$TPR = \frac{TP}{(TP + FN)}$$

Example: To determine what proportion of the actual sick people were correctly detected by the model.

Macro & Weighted Recall

$$Recall_{positive} = \frac{Correct\ positive\ prediciton}{Total\ correct\ positive\ prediction} = \frac{TP}{TP + FN}$$

$$Recall_{negative} = \frac{Correct\ negative\ prediciton}{Total\ correct\ negative\ prediction} = \frac{TN}{TN + FP}$$

$$Macro\ Recall = \frac{Recall_{positive} + Recall_{negative}}{2}$$

$$Weighted\ Recall = \frac{n_1 \times Recall_{positive} + n_2 \times Recall_{negative}}{n_1 + n_2}$$

n_1 is number of positive class and n_2 is number of support examples for negative class.

Specificity (True Negative Rate)

It corresponds to the proportion of negative data points that are correctly considered as negative, with respect to all negative data points.

$$TNR = \frac{TN}{(TN + FP)}$$

False Positive Rate: $FP / (FP+TN)$.

False Positive Rate corresponds to the proportion of negative data points that are mistakenly considered as positive, with respect to all negative data points.

False Negative Rate

$$FNR = \frac{FN}{TP + FN}$$

FNR tells us what proportion of the positive class got incorrectly classified by the classifier.

A higher TPR and a lower FNR is desirable since we want to correctly classify the positive class.

F Score

It combines the precision and recall of the model, and it is defined as the harmonic mean of the model's precision and recall.

Harmonic mean between two numbers a and b is defined as

$$\mathbf{H = 2/\,(1/a + 1/b)}$$

$$H = 2ab/(a+b)$$

$$F_\beta = (1+\beta^2) \times \frac{\text{precision} \times \text{recall}}{(\beta^2 \times \text{precision}) + \text{recall}}$$

- Take β > 1 if we want to give more weightage to recall.
- Take β <1 if want to give more weightage to precision.

If β =1 then

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Macro & Weighted F_1 Score

$$F1\ score\_positive = \frac{2 * Precsion_{positive} * Recall_{positve}}{Precsion_{positive} + Recall_{positve}}$$

$$F1\ score\_negative = \frac{2 * Precsion_{negative} * Recall_{negative}}{Precsion_{negative} + Recall_{negative}}$$

$$Macro\ F1\ score = \frac{F1\ score\_positive + F1\ score\_negative}{2}$$

$$Weighted\ Precision = \frac{n_1 \times F1\ score\_positive + n_2 \times F1\ score\_negative}{n_1 + n_2}$$

Here n_1 is number of positive class and n_2 is number of support examples for negative class.

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

Confusion Matrix for Multi-Class Classification

Multiclass classification

Predicted →

Actual ↓	Setosa	Versicolor	Virginica
Setosa	V_{11}	V_{12}	V_{13}
Versicolor	V_{21}	V_{22}	V_{23}
Virginica	V_{31}	V_{32}	V_{33}

Setosa

$TP = V_{11}$

$FP = V_{21} + V_{31}$

$TN = V_{22} + V_{23} + V_{32} + V_{33}$

$FN = V_{12} + V_{13}$

Find all four for the remaining classes.

Confusion Matrix for Multi-Class Classification

		Predicted			
		Iris-setosa	Iris-versicolor	Iris-virginica	Actual Total
Actual	Iris-setosa	19	0	0	19
	Iris-versicolor	0	15	0	15
	Iris-virginica	0	1	15	16
	Predicted Total	19	16	15	50

Thanks