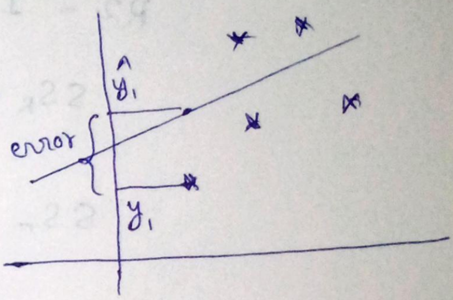
13-Nov-2023

Regression Metrics

1. MAE – Mean Absolute Error
2. MSE – Mean Squared Error
3. RMSE – Root mean Squared Error
4. R2 Score – also called coefficient of determination
5. Adjusted R2-Score

1. Mean Absolute Error:

$$MAE = \frac{|y_1 - \hat{y}_1| + |y_2 - \hat{y}_2| + \cdots |y_n - \hat{y}_n|}{n}$$



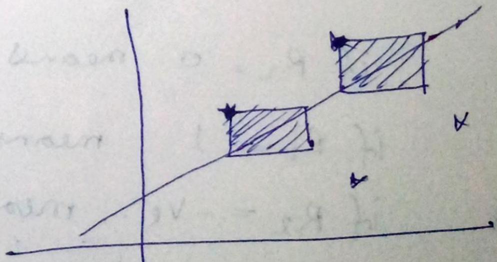
$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}$$

advantage
- same unit
- Roburst outliers

disadvantage
- not differentiable at zero

2. Mean Square Error:

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}$$



advantage:
- use as loss function
- differentiable.

disadvantage
- the squared the unit
- penalize the outliers / not Roburst to outliers

3. RMSE: (Root mean squared Error):

advantage -
- same unit

$$Rmse = \sqrt{MSE}$$

$$= \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

disadvantages:
- not robust to outliers

4. R2 - Score:

also called coefficient of determination and goodness of fit

$$R2 = 1 - \frac{SS_R}{SS_m}$$

SS_R = sum of squared error in regression line

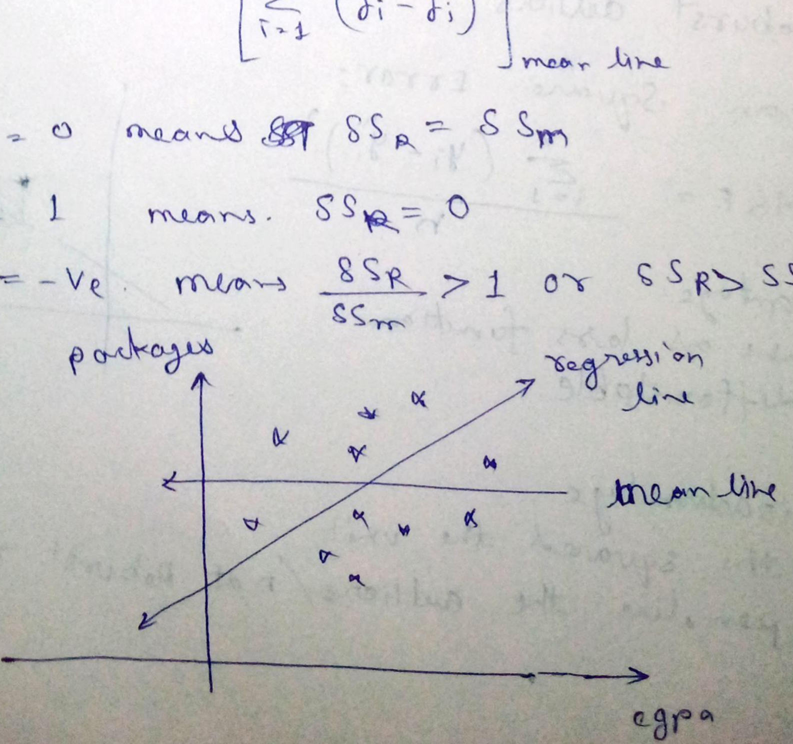
SS_m = sum of squared error in mean line.

$$R_2 = 1 - \frac{\left[\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right]_{\text{Regression line}}}{\left[\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right]_{\text{mean line}}}$$

if $R_2 = 0$ means $SS_R = SS_m$

if $R_2 = 1$ means. $SS_R = 0$

if $R_2 = -ve$ means $\frac{SS_R}{SS_m} > 1$ or $SS_R > SS_m$



5. Adjusted R2-Score:

$$R2(adjusted) = 1 - \left[\frac{(1-R_2)(n-1)}{(n-1-k)} \right]$$

* if irrelavant R2-score is added then adjusted R2-score is decrease
* if relevant column is added then adjusted R2-score is increase.
* adjusted R2-score is useful when we have a multiple linear Regression.

18 March 2024

Classification Metrics

classification Metrics
- accuracy
- confusion
- precesion
- Recall
- F1-Score
- ROC
- AUC

1. Accuracy:

$$\text{accuracy} = \frac{\text{No. of correct prediction}}{\text{Total no. of prediction}}$$

intervies question:
- How much accuracy is good?

Answer - it is based on problem and data

* The accuracy doesn't explain what is the type of गलती for example 90% accuracy it means 10% गलती

2. Confusion Matrix

		prediction 1	prediction 0	
Actual	1	True positive TP	False Negative FN	→ Type-2 error
	0	False positive FP	true Negative TN	
		→ type-1 error		

- using confusion matrix we can calculate the accuracy but ~~one~~ reverse is not true.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

accuracy and confusion is not effective for ~~em~~ imbalanced dataset.

(3.) Precision

A $\quad P_A$

	Sent to spam	Not Sent to spam
Spam	100	170
Not spam	30	700

model-1

B $\quad P_B$

	Sent to spam	Not sent to spam
spam	100	1900
not spam	10	700

model-1

in both the model the accuracy is same. 80% using this we can't select the best one. but

~~PA PPP~~ $FP_A > FP_B$ and $FN_A < FN_B$

हम उस model को select करेंगे जो FP वाली गलती कम करेगा।

Hence in above best model is model-2

precision definition:
what proportion of predicted positives is truly positive.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$P_B > P_A$
select the B

4.) Recall

	Detected cancer	Not Detected
Has cancer	1000	200
No cancer	800	8000

model-A

	Detected cancer	Not Detected
Has cancer	1000	500
No cancer	500	8000

model-B

$$\text{Recall} = \frac{TP}{TP + FN}$$

what proportion of actual positives is correctly classified?

$R_A > R_B$ hence select model-A

5.) F1-Score:

it is combination of precision and recall.

$$F1\ score = \frac{2PR}{P+R}$$ where P = Precision, R = Recall

Here we used Harmonic mean.

Multi-class Precision and Recall:

(i) Precision

	Dog	Cat	Rabbit	Total
Dog	25	5	10	40
Cat	0	30	4	34
Rabbit	4	10	20	34
total	29	45	34	

$\text{Precision (Dog)} = \frac{25}{29} = 0.86$ $\text{Precision (cat)} = \frac{30}{45} = 0.66$

$$\text{Precision (Rabbit)} = \frac{20}{34} = 0.58$$

combined precision → macro technique

$$\frac{0.86 + 0.66 + 0.58}{3} = 0.70$$

→ weighted precision techniques

$$\frac{40}{108} \times 0.86 + \frac{34}{108} \times 0.66 + \frac{34}{108} + 0.58 = 0.71$$

* if all classes in data is balanced then we use macro precision else weighted precision

(ii) Recall

$$\text{Recall (Dog)} = \frac{25}{40} = 0.62; \quad \text{Recall (cat)} = \frac{30}{34} = \quad \text{Recall(Rabbit)} = \frac{20}{34} = 0.58$$

Now we can calculate the macro Recall or weighted Recall

(iii) F1-Score

Same process as above example

$$\text{F1-score (Dog)} = \frac{2\,\text{Precision (Dog)} \cdot \text{Recall (Dog)}}{\text{Precision (Dog)} + \text{Recall (Dog)}}$$

Same for other ~~an~~ animal class

Now all these calculation them you can calculate the macro and weighted F1-score.

* for calculate all these in python using Sklearn :

```
from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred1))
```