WELCOME TO

TECH I.S.

Performance Metrics

Metrices

Until now, we relied on the cost function value to be minimum for regression and classification, but there are other metrics can be used to better evaluate and understand the model.

For Classification: Accuracy, Precision, Recall, F1-score, ROC curves can be used.

For Regression: Normalized RMSE, Normalized Mean Absolute Error (NMAE) can be used.

Accuracy

Accuracy is a measure of how close a given set of guessing from our model are closed to their true value:

- Accuracy is a measure of how well a binary classifier correctly identifies or excludes a condition.
    - It's the proportion of correct predictions among the total number of cases examined.

$$\text{Accuracy} = \frac{\#\ \text{Correct classifications}}{\#\ \text{All classifications}}$$

Disease Classification - Skewed classes

Consider Training logistic regression model h(x) , with y = 1 if disease, y = 0 otherwise.

- The y = 1 class has very few samples as compared to the y = 0 class.
- We got 1% error on test set (99% correct diagnoses)
- Only 0.5% of patients actually have disease.
- If we use this classifier then we get 99.5% of accuracy but only when class = 0 BUT for class y=1 we cannot be sure

For skewed classes, the accuracy metric can be deceptive.

Precision and Recall

<u>Suppose we want to detect for class y = 1:</u>

- <u>Precision</u>: How much we are precise in the detection?
  - Of all patients where we classified y = 1, what fraction actually has the disease?
- <u>Recall</u>: How much we are good at detecting?
  - Of all patients that actually have the disease, what fraction did we correctly detect as having the disease?

Confusion matrix

Actual class

Estiamted class		1 (p)	0 (n)
	1 (Y)	True positive (TP)	False positive (FP)
	0 (N)	False negative (FN)	True negative (TN)

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}$$

$$Precision = \frac{T_p}{T_p + F_p}$$

$$Recall = \frac{T_p}{T_p + T_n}$$

F1- score

It is usually better to compare models by means of one number only.

The **F1-score** can be used to combine precision and recall by taking harmonic mean of them.

$$F_1\text{score} = 2\frac{P \cdot R}{P + R}$$

	Precision(P)	Recall (R)	Average	F_1 Score	
Algorithm 1	0.5	0.4	0.45	0.444	The best is Algorithm 1
Algorithm 2	0.7	0.1	0.4	0.175	
Algorithm 3	0.02	1.0	0.51	0.0392	

Algorithm 3 classifies always 1

Average says not correctly that Algorithm 3 is the best

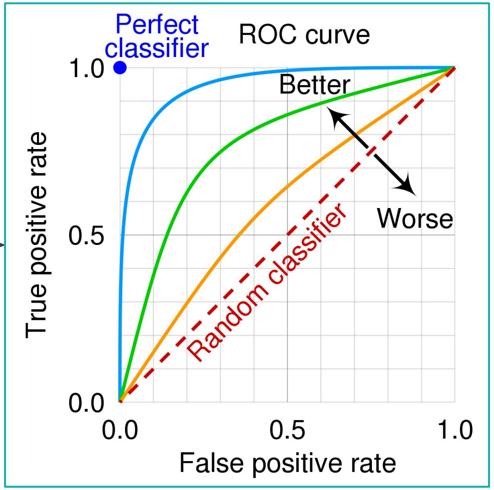
ROC Curves

ROC curves are used to compare the performance of multiple classification models.

ROC curves represents the probability that a randomly chosen positive instance will be ranked ahead of randomly chosen negative instance.

- **(0,0) : classify always negative.**
- **(1,1) : classify always positive.**

- **Diagonal line: random classifier.**
- **Below diagonal line: worse than random classifier Different classifiers can be compared.**



Much obliged.

TECH I.S.