Department of AI&DS

MACHINE LEARNING
22AD2203R

Topic:

PERFORMANCE METRICS

Session - 03

Dr. NARENDRA BABU TATINI
Associate Professor
Department of IoT

AIM OF THE SESSION

To build an accurate and efficient machine learning model that can handle both classification and regression tasks.

INSTRUCTIONAL OBJECTIVES

This session is designed to:

1. Understand the metrics to monitor and measure the performance of a machine learning model.
2. Apply metrics to solve classification and regression problems.

LEARNING OUTCOMES

At the end of this session, you should be able to:

1. Describe the different metrics used to monitor and measure the performance of a machine learning model, and
2. Apply metrics to validate the performance of output generated by a machine learning model.

PERFORMANCE METRICS

- How to validate the performance of output generated by a machine learning model?
- Metrics are needed to monitor and measure the performance of a model.
- In general, machine learning problems have been divided into regression and classification problems.
- Hence, metrics are divided into:
  - Regression metrics
  - Classification metrics.

REGRESSION METRICS

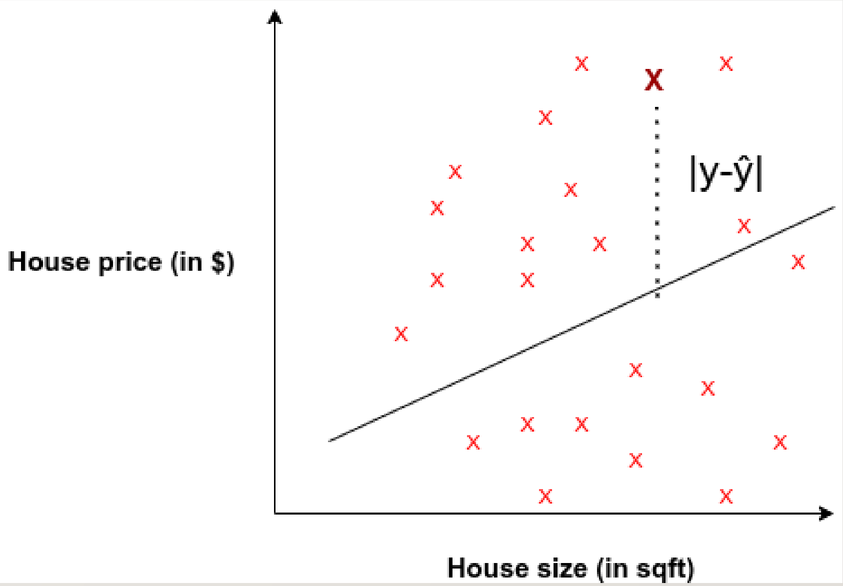
- Regression models generate continuous output.
- Hence, a distance-based calculation between the predicted output and ground truth data is essential.
- The most popular metrics to evaluate the regression models are:
    - Mean Absolute Error (MAE),
    - Mean Squared Error (MSE),
    - Root Mean Squared Error (RMSE),
    - R^2 (R-Squared),
    - Adjusted R^2.

REGRESSION METRICS

- **Mean Absolute Error (MAE)**

- Mean Absolute Error is the average of the difference between the ground truth and the predicted values.
- Mathematically, it is represented as :

$$MAE = \frac{1}{N}\sum_{J=1}^{N}\left|y_j - \hat{y}_j\right|$$



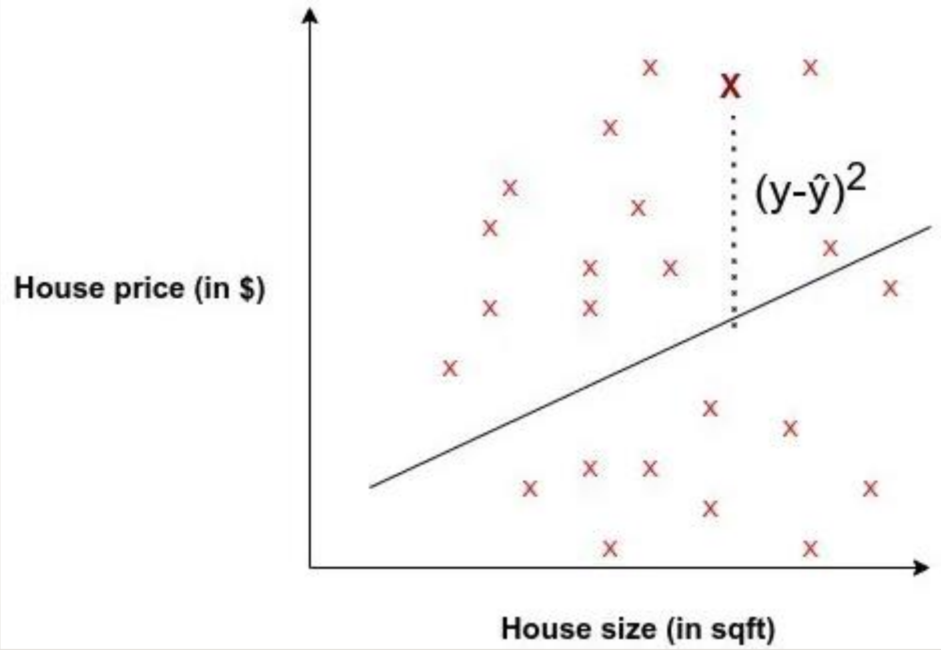
- It's more robust towards outliers.
- Error interpretation needs no second thoughts.
- It gives us a measure of how far the predictions were from the actual output.

REGRESSION METRICS

- **Mean Squared Error (MSE)**

- Mean Squared Error is the average of the squared difference between the ground truth and the value predicted by the regression model.
- Mathematically, it is represented as :

$$MSE = \frac{1}{N}\sum_{J=1}^{N}\left(y_j - \hat{y}_j\right)^2$$



- It's more prone to outliers than other metrics.
- It is differentiable; hence it can be optimized better.

REGRESSION METRICS

- **Root Mean Squared Error (RMSE)**

- Root Mean Squared Error is the square root of the average of the squared difference between the ground truth and the value predicted by the regression model.
- Mathematically, it is represented as :

$$RMSE = \sqrt{\frac{1}{N}\sum_{J=1}^{N}\left(y_j - \hat{y}_j\right)^2}$$

- Error interpretation can be done smoothly.
- It's less prone to outliers.
- It is differentiable; hence it can be optimized better.

REGRESSION METRICS

- **R-Squared (R^2)**

- The R-squared metric enables to comparison of the model with a constant baseline to determine the performance of the regression model.
- Mathematically, it is represented as :

$$R^2 = 1 - \frac{MSE\ (Model)}{MSE\ (Baseline)}$$

- If the sum of the Squared Error of the regression line is small, R^2 will be close to 1 (Ideal), meaning the regression was able to capture 100% of the variance in the target variable.

REGRESSION METRICS

- **Adjusted R^2**

- When the model overfits the data, the variance will be 100% but the model learning hasn't happened. To overcome this problem, R^2 is adjusted with the number of independent variables.
- Mathematically, it is represented as :

$$R^2 = 1 - \left[\left(\frac{n-1}{n-k-1}\right) * \left(1 - R^2\right)\right]$$

n = number of observations,

k = number of independent variables

- The adjusted R^2 always lower than R^2.
- It only shows improvement if there is a real improvement.

CLASSIFICATION METRICS

- Classification models generate discrete output.
- Hence, a metric is required that compares discrete classes.
- The most popular metrics to evaluate the classification models are:
  - Confusion Matrix,
  - Precision and Recall,
  - F1-score,
  - AU-ROC,
  - Accuracy.

CLASSIFICATION METRICS

- **Confusion Matrix**

- Confusion Matrix is the easiest way to measure the performance of the classification model.
- TP signifies how many positive class samples your model predicted correctly.
- TN signifies how many negative class samples your model predicted correctly.
- FP signifies how many negative class samples your model predicted incorrectly.
- FN signifies how many positive class samples your model predicted incorrectly.

		True value	
		1	**0**
Predicted value	**1**	True Positive (TP)	False Positive (FP)
	0	False Negative (FN)	True Negative (TN)

CLASSIFICATION METRICS

- **Precision and Recall**

- Precision is defined as the ratio of TP to the total number of predictions as positives.
- Mathematically, it is represented as :

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

- Recall is defined as the ratio of TP to the total number of actual positives.
- Mathematically, it is represented as :

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

CLASSIFICATION METRICS

- **F1-score**

- F1-score is the harmonic mean of precision and recall.
- Mathematically, it is represented as :

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

- It gives equal importance to precision and recall.
- It presents a good balance between precision and recall and gives good results on imbalanced classification problems.

CLASSIFICATION METRICS

- **AU-ROC (Area Under Receiver Operating Characteristics Curve)**

- AU-ROC makes use of True Positive Rates (TPR) and False Positive Rates (FPR) to visualize the performance of the classification model.
- Mathematically, it is represented as :

$$\boldsymbol{TPR = \frac{TP}{TP + FN}} \qquad \boldsymbol{FPR = \frac{FP}{FP + TN}}$$

- High ROC means that the probability of a randomly chosen positive example is indeed positive.
- ROC curves aren't a good choice when your problem has a huge class imbalance.

CLASSIFICATION METRICS

- **Accuracy**

- Accuracy tells the overall effectiveness of the classifier.
- Mathematically, it is represented as :

$$Accuracy = \frac{TP + TN}{N}$$

N is the total sample size.

- It is the simplest metric to use and implement.

Self-Assessment Questions

1. Which among the following evaluation metrics would you NOT use to measure the performance of a classification model?

(a) Precision
(b) Recall
(c) Mean Squared Error
(d) F1-score

2. The true-positive rate is also referred to as

(a) Recall
(b) Accuracy
(c) Precision
(d) F1-score

Self-Assessment Questions

3. A single metric which combines both precision and recall is the

(a) Precision
(b) Recall
(c) Mean Squared Error
(d) F1-score

4. What is called the average squared difference between classifier predicted output and actual output?

(a) Mean Squared Error
(b) Mean Absolute Error
(c) Root Mean Squared Error
(d) Mean Relative Error

REFERENCES FOR FURTHER LEARNING OF THE SESSION

Text Books:

1. Mitchell, Tom. Machine Learning. New York, NY: McGraw-Hill, 1997. ISBN: 9780070428072.
2. MacKay, David. Information Theory, Inference, and Learning Algorithms. Cambridge, UK: Cambridge University Press, 2003. ISBN: 9780521642989.

Reference Books:

1. EthemAlpaydin “Introduction to Machine Learning “, The MIT Press (2010).
2. Stephen Marsland, “Machine Learning an Algorithmic Perspective” CRC Press, (2009).

Sites and Web links:

1. Data Science and Machine Learning: https://www.edx.org/course/data-science-machinelearning.
2. Machine Learning: https://www.ocw.mit.edu/courses/6-867-machine-learning-fall-2006/.

THANK YOU

Team – MACHINE LEARNING