School of Computer Science Engineering and Technology

Course- BTech
Course Code: CSET 301
Year- 2022-23
Date-

Type- Specialization Core
Course Name- AIML
Semester- 4th
Batch-

Lab Assignment W4L1

Classification metrics

Choosing right evaluation metrics for the problem is one of the most important aspects of machine learning. Choice of metrics allows us to compare performance of different models and helps in model selection.

In this task, we will explore following metrics:

confusion matrix

accuracy

precision

recall

f1 score

Dataset

The training dataset is available at "data/ozone_levels_train.csv" in the respective challenge' repo.

The testing dataset is available at "data/ozone_levels_test.csv" in the respective challenge' repo.

The dataset is modified version of the dataset 'ozone level' on provided by UCI Machine Learning repository.

Original dataset: https://archive.ics.uci.edu/ml/datasets/Ozone+Level+Detection

Objective

To learn about classification metrics and compare logistic regression and decision tree on the same dataset

Tasks

define X(input) and Y(output)

train the decision tree model

train the logistic model

construct a confusion matrix

calculate the classification accuracy

calculate the Precision

calculate the Recall

calculate the F1 score

calculate Area Under ROC Curve

Further fun

Calculate precision and recall

find the area under the curve for Roc metrics

Implement below metrics using inbuilt libraries

confusion matrix

accuracy

precision

recall

f1 score

Helpful links

Classification metrics with google developers: https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative

classification metrics: https://www.kdnuggets.com/2020/04/performance-evaluation-metrics-classification.html

pd.get_dummies() and One Hot Encoding: https://queirozf.com/entries/one-hot-encoding-a-feature-on-a-pandas-dataframe-an-example

Differences between Logistic Regression and a Decision Tree: https://www.geeksforgeeks.org/ml-logistic-regression-v-s-decision-tree-classification/

Decision Tree Classifier by https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html

Understanding classification metrics like Precision, Recall, F-Scores and Confusion matrices:

https://nillsf.com/index.php/2020/05/23/confusion-matrix-accuracy-recall-precision-false-positive-rate-and-f-scores-explained/

Understanding the ROC Curve: https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc