XGBoost Regressor in Machine Learning

By Ramani Nagarajan

Agenda

- Overview
- How XGBooster Regressor Works
- Example: XGBooster Regressor in Python

Overview

- XGBooster Regression is a supervised machine learning algorithm based on gradient boosting. It builds an ensemble of decision trees where each tree tries to correct the errors of the previous trees, leading to a highly accurate predictive model.

How XGBooster Regressor Works

Initial Prediction: The model starts with an initial prediction (often the mean of the target values).

Residual Calculation: Calculates residuals (errors) between predicted and actual values.

Decision Tree Construction: Builds a new decision tree to predict these residuals (errors).

Sequential Boosting: Adds new trees in sequence; each learns from previous mistakes and is weighted according to its contribution.

Ensemble Output: The final prediction is computed as a weighted sum of the predictions from all trees.

Loss Function Optimization: Both a loss function (like RMSE) and a regularization term are minimized during training to promote accuracy and prevent overfitting.

Parameters: Includes hyperparameters such as max_depth, learning_rate, subsample, and regularization parameters to fine-tune the model.

Example: AdaBooster Regressor in Python

```
import pandas as pd
##Open CSV and assign values
dataset=pd.read_csv("50_Startups.csv")

dataset=pd.get_dummies(data,dtype=int,drop_first=True)
dataset.columns


independent=dataset[['R&D Spend', 'Administration',
'Marketing Spend','State_Florida', 'State_New York']]
dependent=dataset[["Profit"]]

##-SPLIT TRAIN & TEST

from sklearn.model_selection import train_test_split as tts

X_Train,X_Test,Y_Train,Y_Test=tts(independent,dependent,te
st_size=0.30,random_state=0)
```

```
##STANDARDIZATION

import xgboost as xg

regressor = xg.XGBRegressor(objective
='reg:linear',n_estimators = 10, seed = 123)
regressor=regressor.fit(X_Train,Y_Train)



##EVALUATION
y_predict = regressor.predict(X_Test)
#R-SQUARE
from sklearn.metrics import r2_score
r_score = r2_score(Y_Test,y_predict)

r_score
```

Thank you