Python Hyperparameter Tuning

WEEK 6

Data Science Academy

Disclaimer: This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset.

Dataset description: It has 768 rows of data.

1. Pregnancies– Number of times pregnant
2. Glucose– Plasma glucose concentration a 2 hours in an oral glucose tolerance test
3. BloodPressure– Diastolic blood pressure (mm Hg)
4. SkinThickness– Triceps skinfold thickness (mm)
5. Insulin- 2-Hour serum insulin (mu U/ml)
6. BMI – Body mass index (weight in kg/ (height in m)^2)
7. DiabetesPedigreeFunction– Diabetes pedigree function
8. Age- Age (years)
9. Outcome – Class variable (0 or 1) 268 of 768 are 1, the others are 0

Task: You will use the Pima Indian diabetes dataset. The dataset corresponds to a classification problem on which you need to make predictions on the basis of whether a person is to suffer diabetes given the 8 features in the dataset. You are asked to use this data to build a XGBoost Classifier model using scikit-learn. Tune the parameters using Grid Search and Randomized Search strategies.

Data Science
Academy