Python – Machine Learning - Functions to Remember

from sklearn.tree import DecisionTreeRegressor

model = DecisionTreeRegressor(random\_state = 1)

model.fit(X, y)

* X is the data, y is the output from the data. Creates a model called model which can be used to predict further data.

Predictions = model.predict(X)

* Predicts the output for some Data X, based on the model computed above.

from sklearn.model\_selection import train\_test\_split

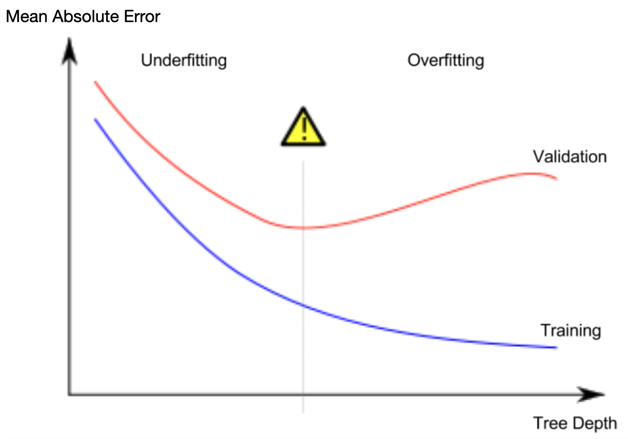
train\_X, val\_X, train\_y, val\_y = train\_test\_split(X, y, random\_state=1)

* train\_test\_split splits the data into 4 categories , the training data and output and the validation data and output. Splits it equally 50-50.

from sklearn.metrics import mean\_absolute\_error

mean\_absolute\_error(val\_y, val\_predictions)

* Gives the mean absolute error between predictions and real values.

- showing difference between overfitting and underfitting.

model = DecisionTreeRegressor(max\_leaf\_nodes=100, random\_state=0)

* This function here inputs max\_leaf, which tell you how many levels deep the machine learning should go, it affects overfitting and underfitting, you can change it to find the optimal absolute mean error to get the best machine learning algorithm