Jan 9th 2021

Better evaluation Using cross-validation:

Two Alternatives:

1. train_test_split(): seperate dataset into training, testing, and validation
   train model against training set and evaluate its performance against validation set.

More accurate:

2. Sklearn: k-fold cross validation feature.
   Split training set into 10 (arbitrary number) subsets.
   → train and evaluate model 10 times:
   → each time train with 9/10 subsets and evaluate with the left subset
   => Result will be an array of validation/evaluation results

Pro: Able to show the preciseness of estimations (standard deviation)

Con: Need to train model multiple times. Not realistic when dataset becomes big.

- X The cross-validation feature from sklearn expects a utility function, instead of a cost function.
  utility function: greater is better → used for showing scores and performance measures
  Cost function: lower is better. → for optimization by minimizing the error score.