

Cross Validation

With GridSearchCV and RandomizedSearchCV

Cross Validation

Cross validation splits data in the given number (let's say 5) of datasets as:

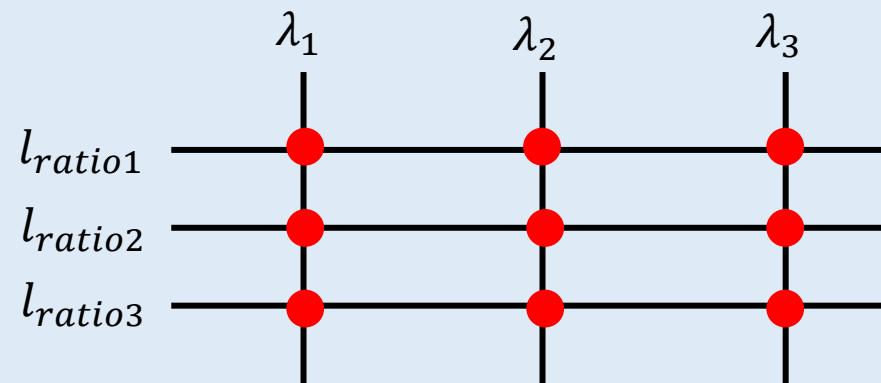
Train	Train	Train	Train	Test
Train	Train	Train	Test	Train
Train	Train	Test	Train	Train
Train	Test	Train	Train	Train
Test	Train	Train	Train	Train

Grid Search Cross Validation (Grid Search CV)

Assume we are finding the best parameters for elastic net regression, there are two hyperparameters we are supposed to take care of

1. λ
2. l_{ratio}

Let's say we have three λ values to try and three l_{ratio} to try. Then for each value of λ there are three l_{ratio} to be tried, so in grid search CV we have



For every combination of l_{ratio} and λ there will be cross validation carried out as per the CV parameter provided to GridSearchCV, let's say cv is 5, then 5 models will be trained for every grid cross red point and average of assessment metric will be recorded for each cross red point on the grid, for above example there will be $9 \times 5 = 45$ models will be trained

Randomised Search CV

In this type of cross validation technique, unlike gridsearchcv where all the combination of hyperparameter is tried out and cross validated, it will randomly select the combination of the parameters and cross validates it and records the best average metric of best parameter combination in the cross validation and best combination of parameters are returned at the end.

When the number of hyperparameters are too many and therefore checking their combinations become really difficult on gridsearchcv as it becomes time consuming and computationally expensive, in such cases RandomisedSearchCV is advised to use since it randomly selects the parameter combination

If the parameters to be tried out have equal importance it is advised to use GridSearchCV