**Summary/Review**

**Cross Validation**

The three most common cross validation approaches are:

* k-fold cross validation
* leave one out cross validation
* stratified cross validation

Cross validation method involves dividing the dataset into 3 parts:

* training set - is a portion of the data used for training the model
* validation set - is a portion of the data used to optimize the hyper-parameters of the model
* test set - is a portion of the data used to evaluate the model

**Cross Validation Syntax**

`Scikit Learn` library contains many methods that can perform the splitting of the data into training, testing and validation sets. The most popular methods that we covered in this module are:

\*   train\_test\_split - creates a single split into train and test sets

\*   K-fold - creates number of k-fold splits, allowing cross validation

\*   cross\_val\_score - evaluates model's score through cross validation

\*   cross\_val\_predict – produces the out-of-bag prediction for each row

\*   GridSearchCV – scans over parameters to select the best hyperparameter set with the best

     out-of-sample score