

10 Fold Cross Validation for Parameter Selection

Cross Validation is the standard method for evaluation in empirical machine learning. It can also be used for parameter selection if we make sure to use the training set only.

To select parameter λ of algorithm $A(\lambda)$ over an enumerated range $\lambda \in [\lambda_1, \dots, \lambda_k]$ using dataset D , we do the following:

1. Split the data D into 10 disjoint folds.
2. For each value of $\lambda \in [\lambda_1, \dots, \lambda_k]$:
 - (a) For $i = 1$ to 10
 - Train $A(\lambda)$ on all folds but i^{th} fold
 - Test on i^{th} fold and record the error on fold i
 - (b) Compute the average performance of λ on the 10 folds.
3. Pick the value of λ with the best average performance

Now, in the above, D only includes the training data and the parameter is chosen without knowledge of the test data. We then re-train on the entire train set D using the chosen parameter value and evaluate the result on the test set.