Cross Validation

Cross Validation

Cross validation is a widely used resampling method. It repeats a calculation multiple times using randomly selected subsets of the complete dataset.

To obtain unbiased estimates of expected model performance while performing model selection, it is necessary to use nested cross validation. As the name implies, nested cross validation is performed though a pair of nested CV loops. The outer loop uses a set of folds to perform model evaluation. The inner loop performs model selection using another randomly sampled set of folds not used for evalution by the outer loop. This algorithm allows model selection and evaluation to proceed with randomly sampled subsets of the full data set, thereby avoiding model selection bias.

In this lab you will perform simple cross validation and nested cross validation.

Lab Steps

1. Make sure that you have completed the setup requirements as described in the Lab Overview section.
2. Now, run jupyter notebook and open the “CrossValidation.ipynb” notebook under Module 5 folder.
3. Examine the notebook and answer the questions along the way.

Question 1

1/1 point (graded)

What is the mean ROC of the cross-validated model?

- 0.707
- 0.792
- 0.773 ✔
- 0.728

Submit You have used 2 of 2 attempts

✔ Correct (1/1 point)

Learn About Verified Certificates