Exercise 2, May 1st

The article by (Rashka et al., 2018) discusses a number of approaches for optimizing the choice of hyperparameters. It is recommended that you read the paper but skip Paragraphs 1.7 (Confidence Intervals via Normal Approximations) and 2.4 (The Bootstrap Method and Empirical Confidence Intervals).

Questions about the text:

1. Explain what is stratified sampling - or stratification - and why it is important.

2. Explain what is pessimistic and optimistic bias and how it relates to the size of the training and test set.

3. What are the three goals of performance estimation?

4. Why do we need a validation set?

5. What is the difference between holdout method, 2-fold cross-validation, and repeated holdout methods?