

Traditional Cross Validation

SCENARIO A: w/ validation and test set (two holdout sets)

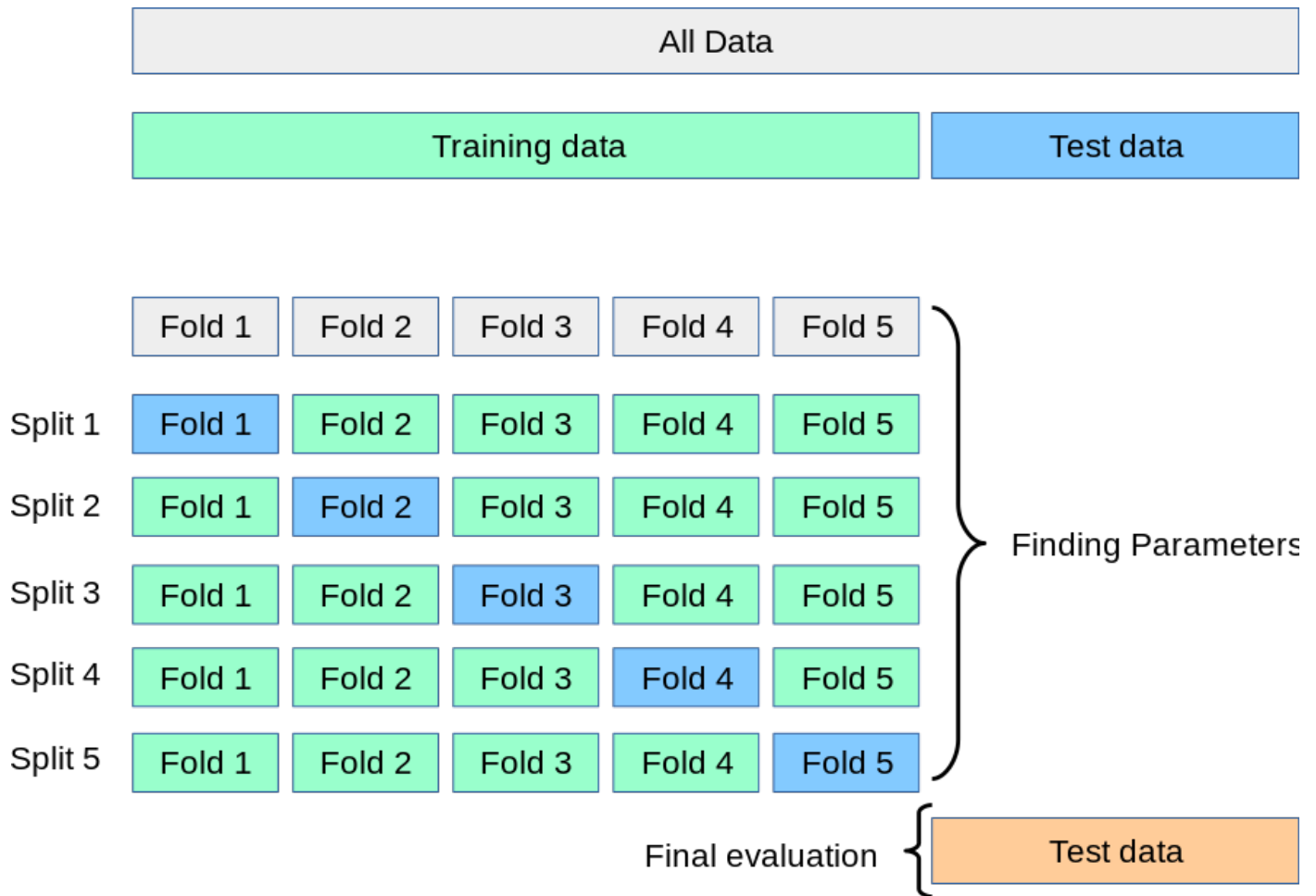
- step₁ randomly split 50% train / 30% validation / 20% test
- step₂ fit model to training set; obtain parameter estimates & evaluate goodness of fit; perform feature selection to determine your best model
- step₃ fit best model to validation set; obtain new parameter estimates & evaluate goodness of fit; compare goodness of fit from step 2 to determine if model is overfit, underfit, or just right.
- step₄ apply best model to test set using the same parameter estimates from step 3; evaluate "out-of-sample" model performance

SCENARIO B: w/ only one holdout set

- step₁ randomly split 70% train / 30% test
- step₂ fit model to training set; obtain parameter estimates & evaluate goodness of fit; perform feature selection to determine your best model
- step₃ fit best model to test set; obtain new parameter estimates & evaluate goodness of fit; compare goodness of fit from step 2 to determine if model is overfit, underfit, or just right.
- step₄ apply best model to test set using the same parameter estimates from step 1; evaluate "out-of-sample" model performance

The only difference between scenario A and scenario B is the number of holdout sets. When you have enough data, scenario A is preferred. In scenario A you have two holdout sets: one for determining if the trained "best" model is overfit or underfit (the "validation set") and a second for determining out of sample performance (the "test" set). In scenario B, the same holdout set serves both purposes. Scenario A is preferred, but scenario B is most common. Note that k-folds cross validation is a little different from either of these scenarios in how it assigns holdout data, but the concept of "out-of-sample" performance is the same. K-folds is a smarter way to handle scenario B...

K-Folds Cross Validation



...K-folds is a smarter way to handle scenario B because it sort of turns the training set into its own validation set as well. K-folds does a great job at preventing overfitting and is used widely in practice. Here again, the precence of "test data" is a luxury not a requirement.