Evaluating Performance II

Lecture 07

Spot the misstep

1

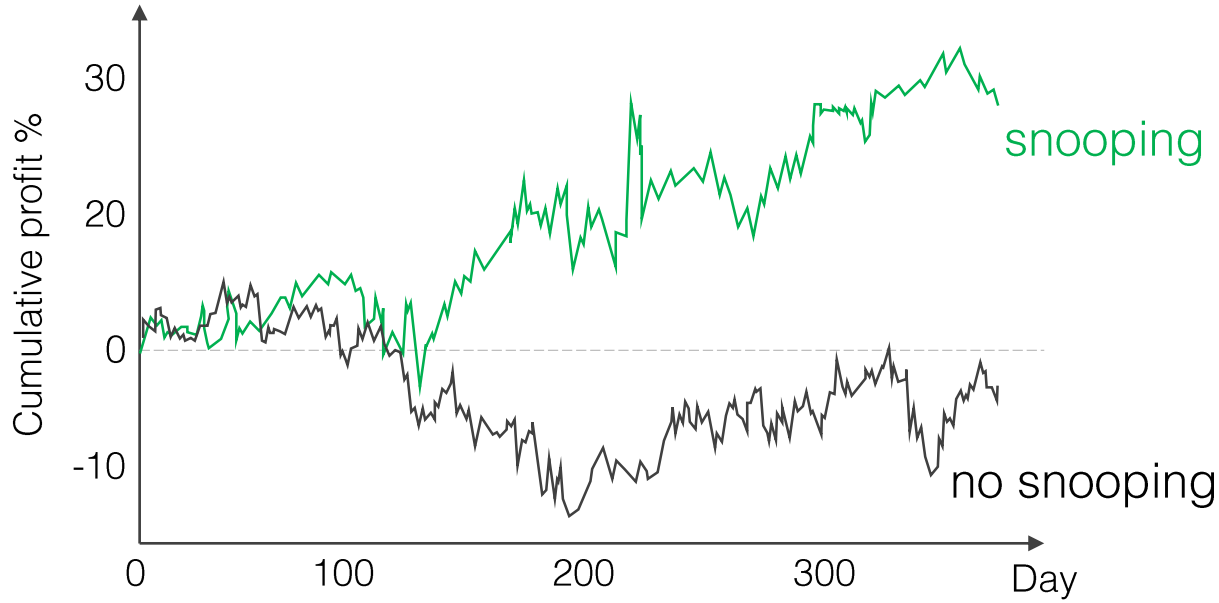
1. Goal: predict the exchange rate for the U.S. Dollar vs British Pound (using 20 past observations)

2. You take your historical data, normalize it, then split it randomly into a training and test set

3. You train on the training data, test on the test data

Results:
Your predictions are correct 56% of the time

Estimate your profits…



Abu-Mostafa, Learning From Data

2

1. Goal: predict the Dow Jones Industrial average

2. You randomly split your data into a training and test dataset

3. Choose a model with lots of flexibility

4. You iterate on the following process dozens of times:
    1. Train your model on the training data
    2. Test your model on the test data
    3. Evaluate performance on the test data

5. Report that you were able to achieve 75% accuracy on your test set!

3

1. Goal: predict long-term performance of a “buy and hold” strategy in stocks

2. You collect 50 years of historical data and include all currently traded companies in the S&P500

3. You randomly split your data into a training and test dataset.

4. You assume you will strictly follow the “buy and hold” strategy

5. You then use apply your model on the current portfolio and predict that you will be rich in retirement!

Abu-Mostafa, Learning From Data

Data snooping

a.k.a. data leakage

If a test data set has affected **any step** in the learning process, its ability to assess the outcome has been **compromised**.

Abu-Mostafa, Learning From Data

Sampling bias

Are the data we're using for machine learning **representative of the population**?

Avoiding data snooping

Don’t touch your test dataset until you’re ready to evaluate your model’s performance

Abu-Mostafa, Learning From Data

Training, Test Split

Learning model parameters

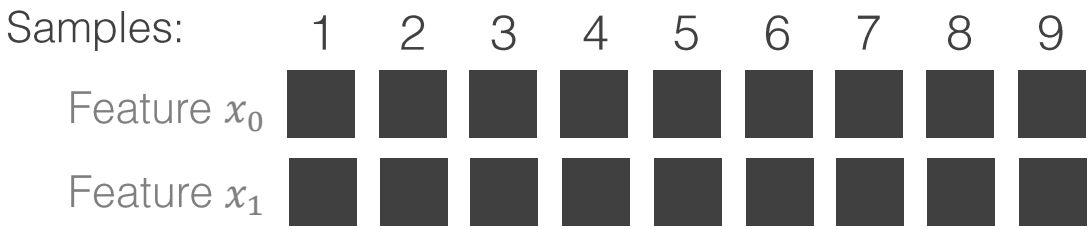
Learn model parameters

Evaluate generalization performance

For small datasets, this reduction in dataset size may be detrimental

Cross-validation

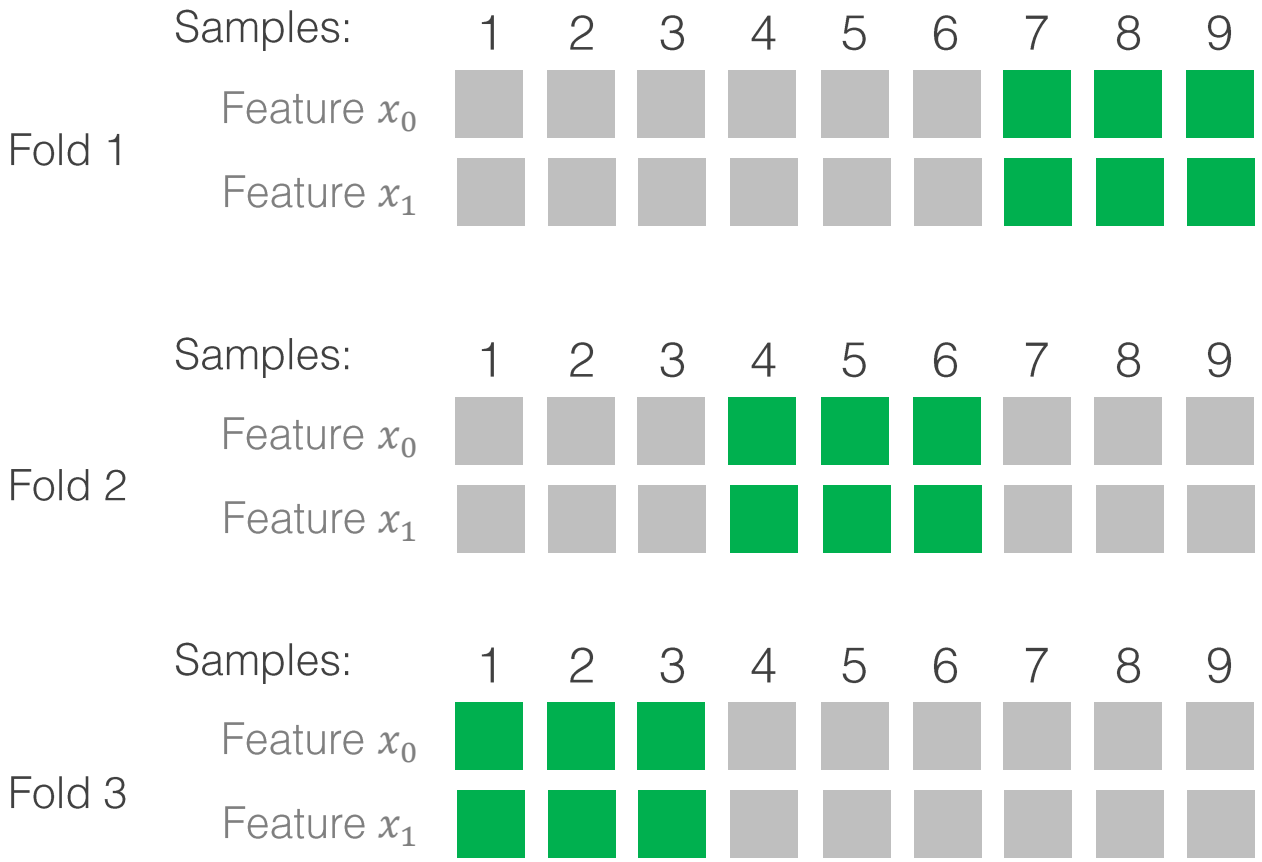
Original feature set with 2 features and 9 samples



Training Data

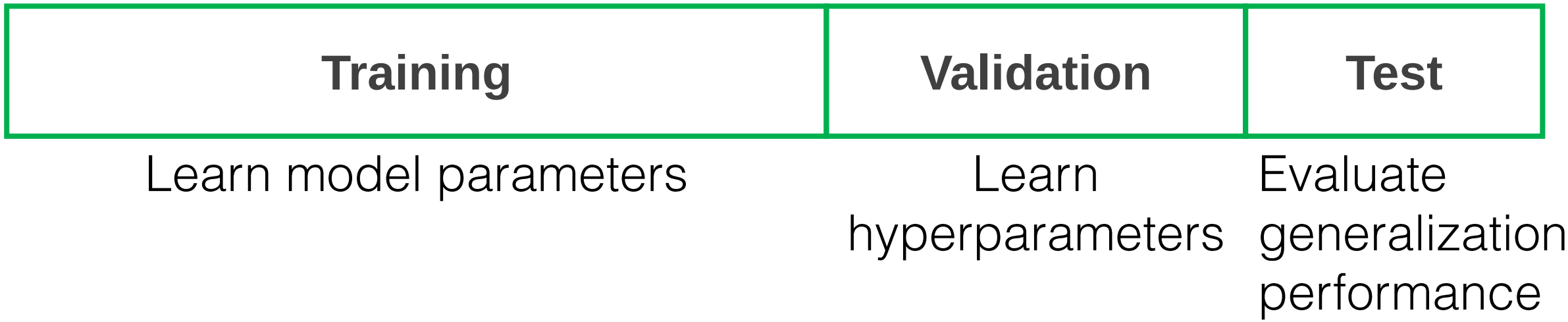
Test Data

K-fold cross validation **K = 3**



Training, Validation, Test Split

Learning parameters AND hyperparameters



Hyperparameters: parameters of your learning algorithm or parameters of you model that are set before training begins

Nested Cross-validation

Instead of a static train/validation/test split, another option is nested cross-validation

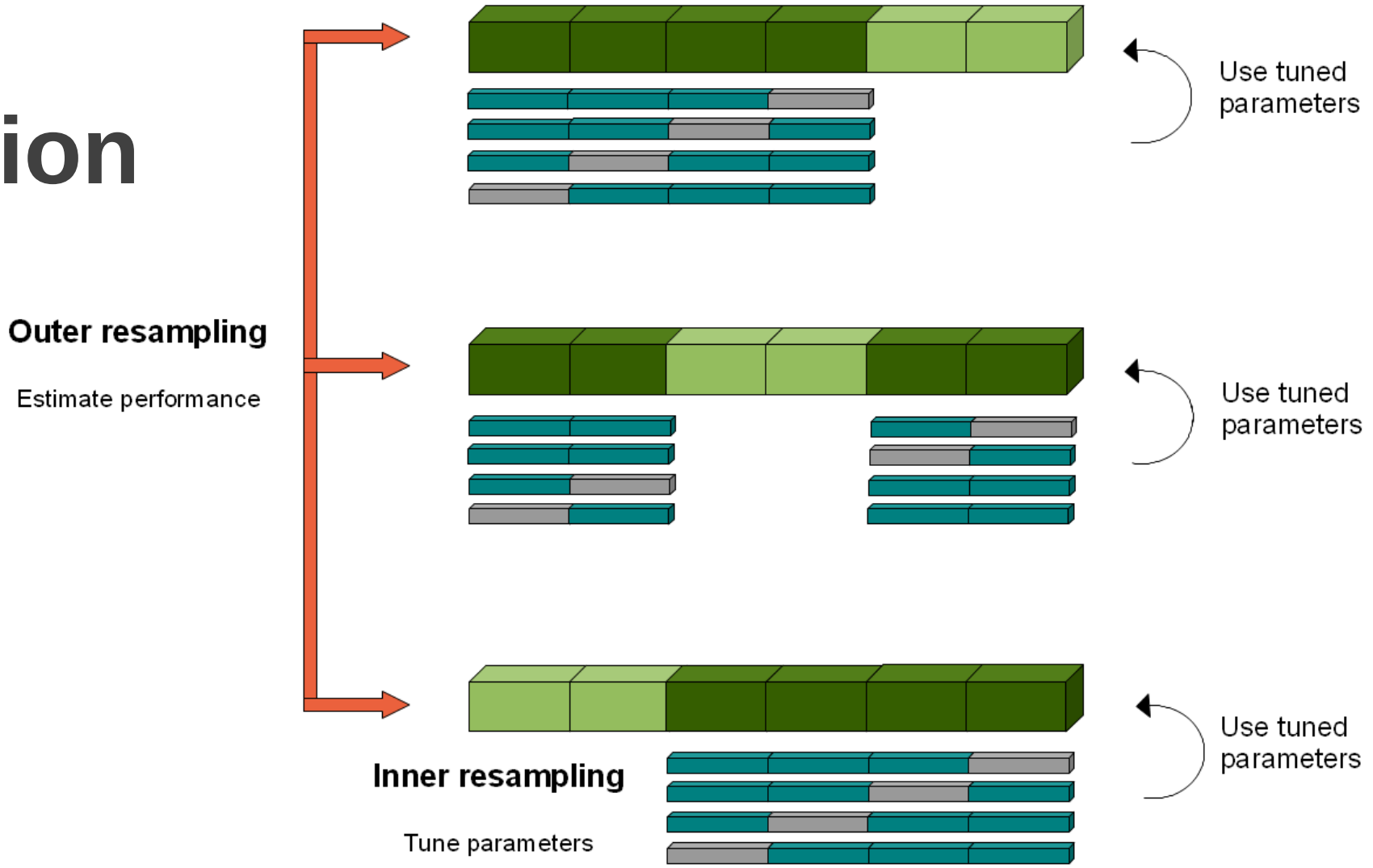


Image Source: https://stats.stackexchange.com/questions/292179/whats-the-meaning-of-nested-resampling

Bootstrap sampling

Sampling **with replacement**

Often used to estimate standard errors and confidence intervals

Integral part of model ensembles (i.e. bagging in random forests)

Abu-Mostafa, Learning From Data