UNIVERSITY OF SAN FRANCISCO
Master of Science in Analytics

Resampling Methods

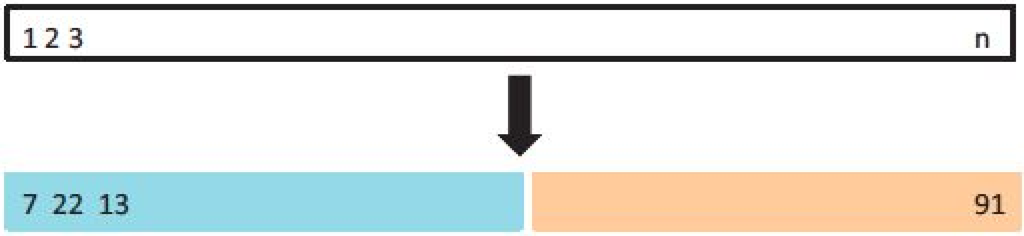
Machine Learning 1

Model assessment & selection

- How do you know whether a model is good?
  - … in a computationally tractable way?
  - … for regression and classification?
- To build a good model, how flexible should it be?
  - What settings should the hyperparameters have?

Is the model good?

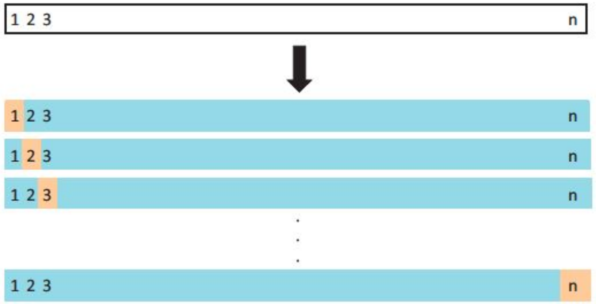
- Performance on test material is lower than on train
  - * Almost always…
  - Why?
- Estimate test performance using training material
  - Randomly-generated portion, *validation set*, not used for training



  - What is the problem with this technique?

Cross-validation

- Types
  - Leave-One-Out Cross-Validation (LOOCV)

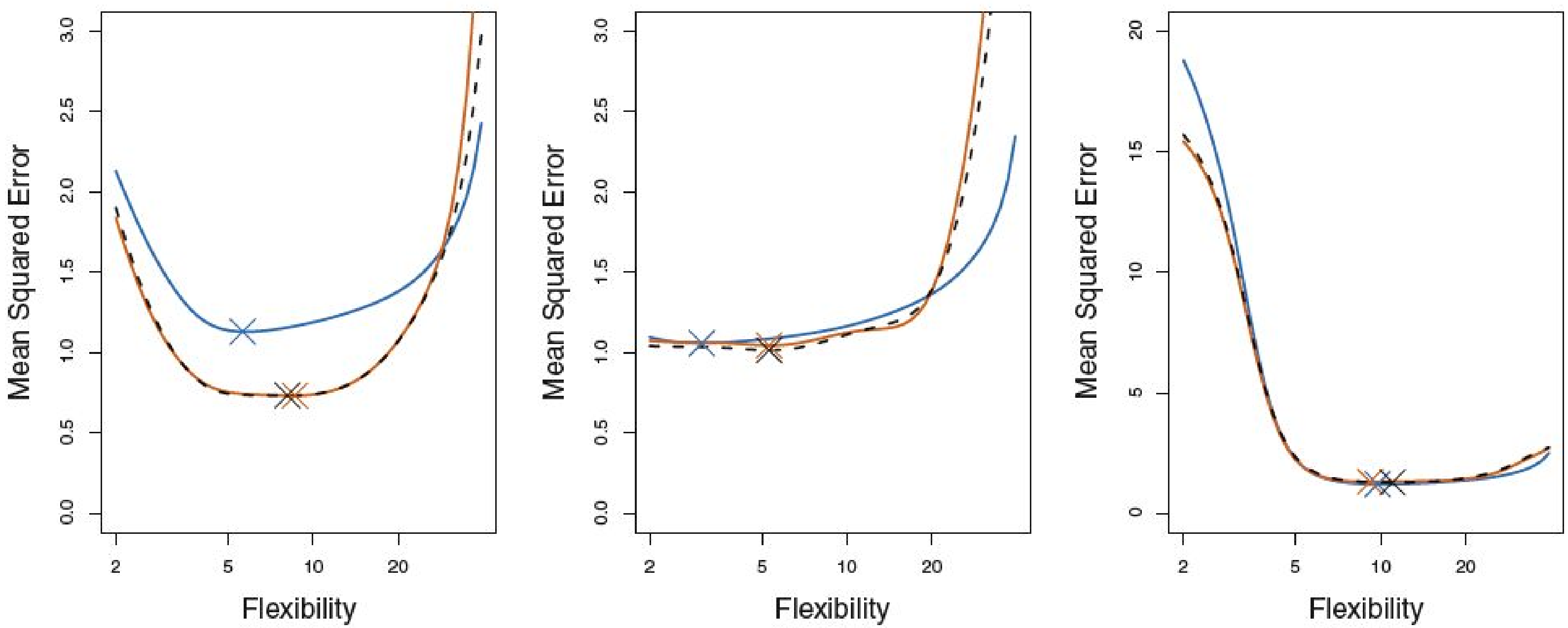


  - k-Fold Cross-Validation
- Results (regression)

$$\mathrm{CV}_{(k)} = \frac{1}{k}\sum_{i=1}^{k}\mathrm{MSE}_i$$

Train vs. test

- Problem: what is the true [read: test] MSE?
    - Examples: LOOCV, k-fold CV and test MSE



    - Curves are similar but suggest different (incorrect) flexibility points
- Can the uncertainty be quantified?

The bootstrap

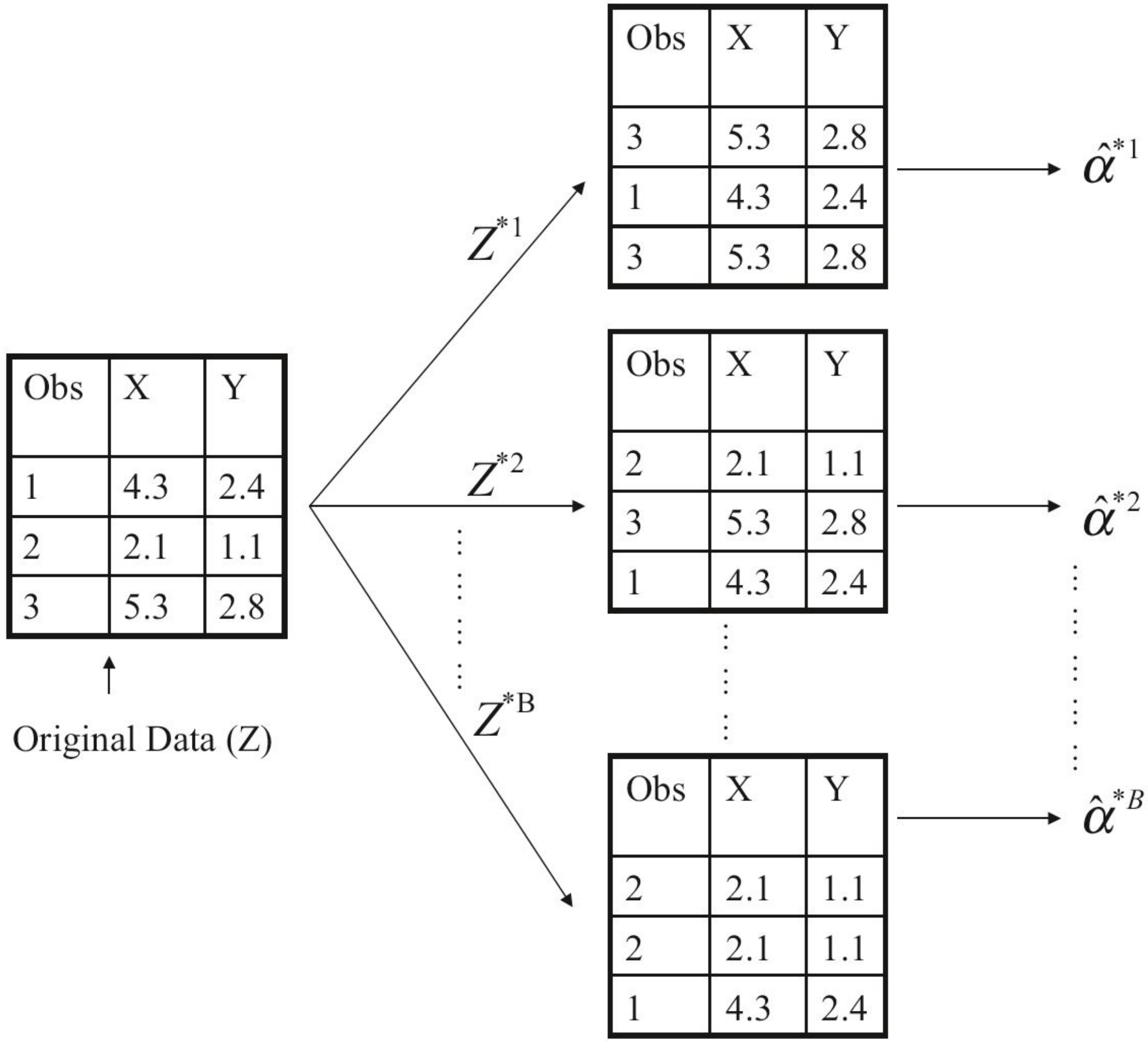
- Objective: quantify uncertainty
    - Sub-objective: decrease variability ("risk")

$$\alpha = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}$$

    - Since all σ terms above are unknown, we estimate from data
    - Curves are similar but suggest different (incorrect) flexibility points
- Algorithm:
    - Choose some (large) value, B (eg. 1000)
    - Create B sets from training by sampling with replacement
    - Compute average (or majority) and SE performance estimates:

$$\bar{\alpha} = \frac{1}{1,000} \sum_{r=1}^{1,000} \hat{\alpha}_r \qquad \mathrm{SE}_B(\hat{\alpha}) = \sqrt{\frac{1}{B-1} \sum_{r=1}^{B} \left(\hat{\alpha}^{*r} - \frac{1}{B} \sum_{r'=1}^{B} \hat{\alpha}^{*r'} \right)^2}$$

Example


Obs X Y
1 4.3 2.4
2 2.1 1.1
3 5.3 2.8
Original Data (Z)
Z^{*1}
Z^{*2}
Z^{*B}
Obs X Y
3 5.3 2.8
1 4.3 2.4
3 5.3 2.8
Obs X Y
2 2.1 1.1
3 5.3 2.8
1 4.3 2.4
Obs X Y
2 2.1 1.1
2 2.1 1.1
1 4.3 2.4
$\hat{\alpha}^{*1}$
$\hat{\alpha}^{*2}$
$\hat{\alpha}^{*B}$