PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY *of* WASHINGTON

Introduction to Machine Learning

MLEARN 510A – Lesson 6

Recap of Lesson 5

- Feature Engineering
- Custom Feature Transformation
- Feature Selection
- Chaining Transformations Together
- Hyper-parameter Tuning
- Testing, Launching, Monitoring and Maintaining

Course Outline

1. Introduction to Statistical Learning
2. Linear Regression
3. Classification
4. Model Building, Part 1
5. Model Building, Part 2
6. **Resampling Methods**
7. Linear Model Selection and Regularization
8. Moving Beyond Linearity
9. Unsupervised Learning
10. Dimensionality Reduction

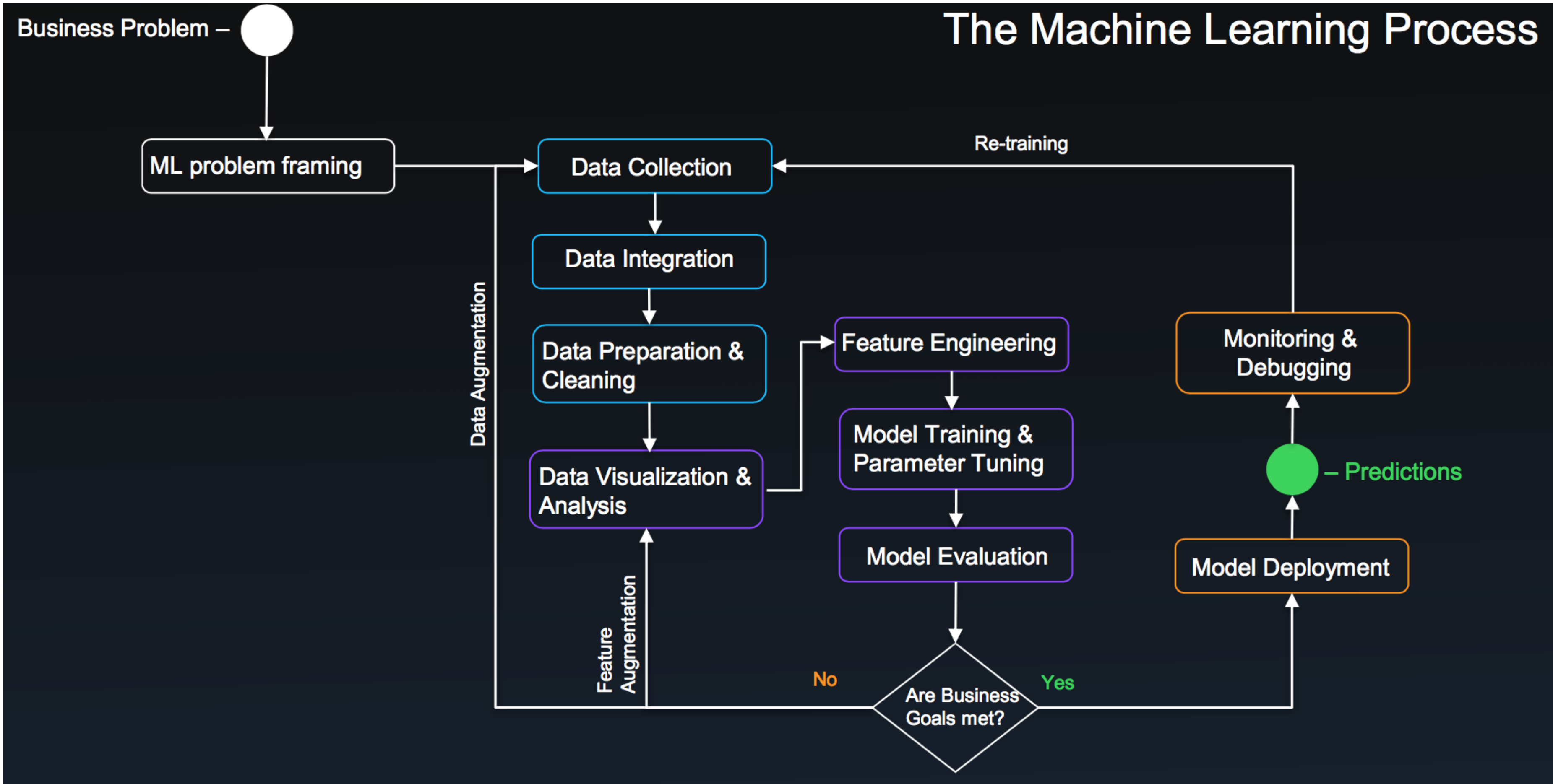
Assignment Solution Review

- Assignments Review

Outline of Lesson 6

- Resampling Methods
- Validation Set Approach
- Leave-One-Out Cross Validation (LOOCV)
- LOOCV vs. k-fold Cross Validation
- Bias-Variance Tradeoff for Cross Validation
- The Bootstrap Method

The Machine Learning Process



Why Resampling?

- **Problem:** Before we deploy our model, we do not have a true test set that can be used to evaluate how well the model generalizes

- **Solution:** Carve out a test set from your training set

Resampling Methods

- Tools that involve <u>repeatedly</u> drawing samples from a training set
- Refit a model of interest on each sample in order to obtain more information about the fitted model
- Useful for
  - Model assessment: Estimate test error rates
  - Model selection: Estimate model flexibility
- **Downside:** They are computationally expensive
- But we have much better computing resources

Types of Resampling Methods

- Two types of resampling methods

- **Cross Validation:** Used to estimate the test error associated with a given statistical learning method in order to evaluate its performance, or to select the appropriate level of flexibility

- **Bootstrap:** provide a measure of accuracy of a parameter estimate or of a given statistical learning method

A Note on Prediction Error

- **Training Error Rate:** How well does the model fits training data

- Training error usually underestimates the **test error** especially if the model is complex

- In general, larger the sample size, lower is the generalization error

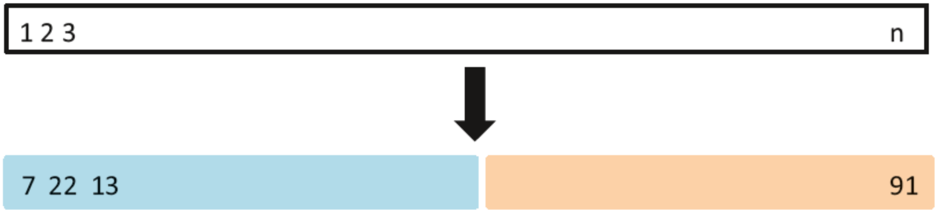
Validation Set/Hold Out Set Approach

- Involves randomly dividing the available set of observations into two parts, a *training set* and a *validation set* or *hold-out set*

- Model is fit on the training set, and the fitted model is used to predict the responses for the observations in the validation set

- The resulting validation provides an estimate of the test error rate

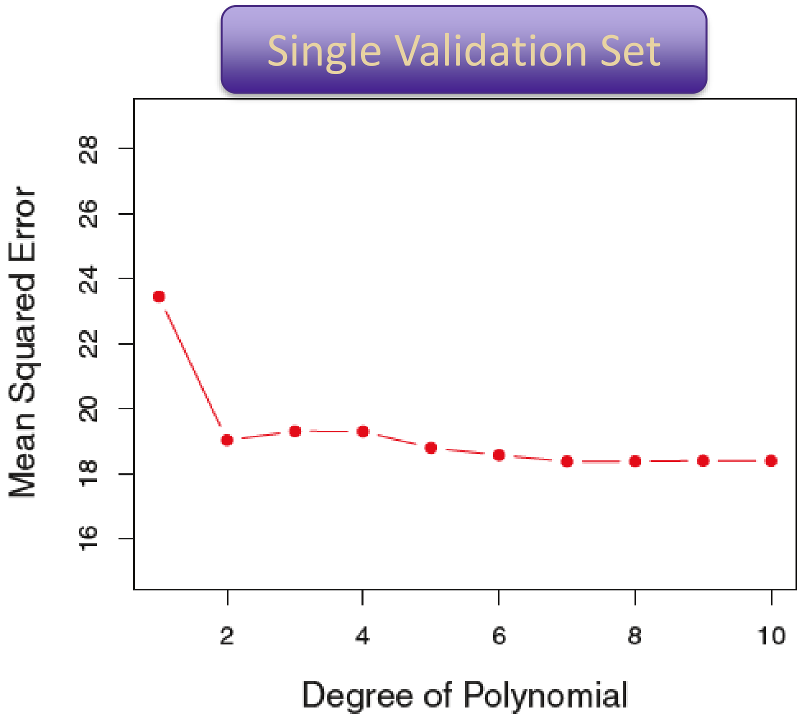
- Simple to implement

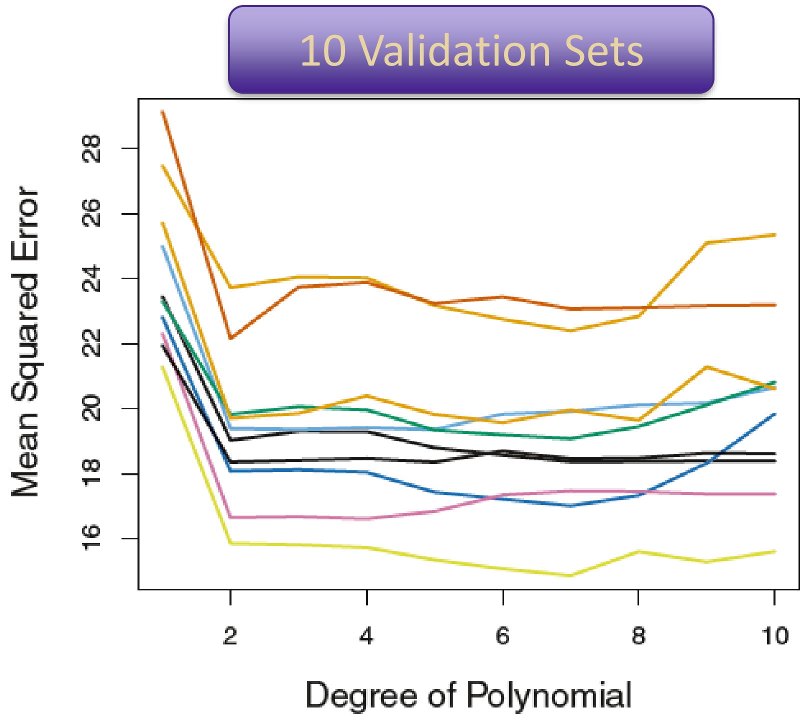


Drawbacks of Hold Out Strategy

- Test error estimated by using a single validation set is high variable
- It highly depends on which observations are included in training set and the test set
- Since training is completed on a smaller set of observations, some statistical models may fit poorly on this smaller set
- In this case, validation set error rate may overestimate the test error for the model fit on entire data

Drawbacks of Hold Out Strategy





Leave-One-Out Cross Validation (LOOCV)

- LOOCV involves splitting the set of observations into two parts
- A single observation (x_1, y_1) is used for the validation set, and the observations $\{(x_2, y_2), . . . , (x_n, y_n)\}$ make up the training set
- Model is fit on the $n - 1$ training observations, and a prediction $\hat{y}_1$ is made for the excluded observation, using its value x_1
- This provides an unbiased estimate of the true error

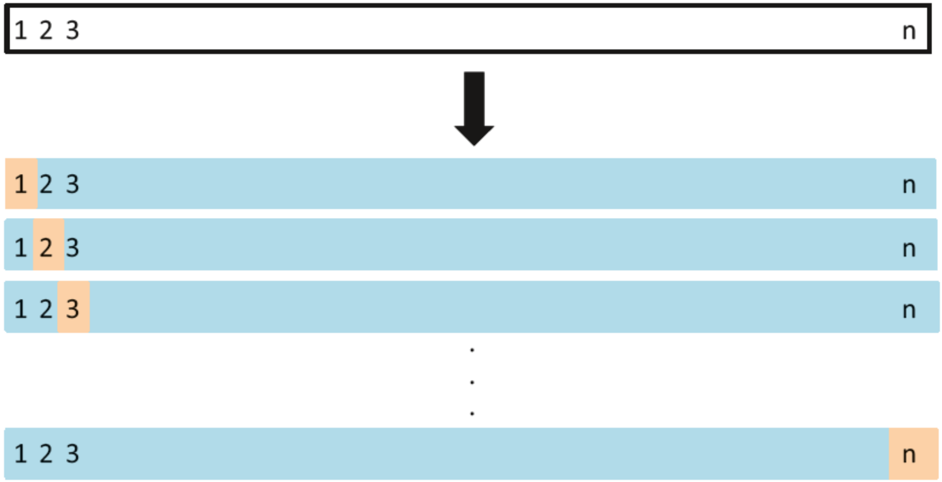
Quiz

- What could go wrong with an error estimate that is based on a single observation in the test set?

Leave-One-Out Cross Validation (LOOCV)

- Basing error estimate on a single observation will have a lot of variability
- Cycle through observations and iteratively include each observation in the test set and measure its error
- LOOCV estimate of the test MSE is

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



Drawbacks of LOOCV

- LOOCV requires a model be fit for each observation in the test set

- This makes it expensive to implement

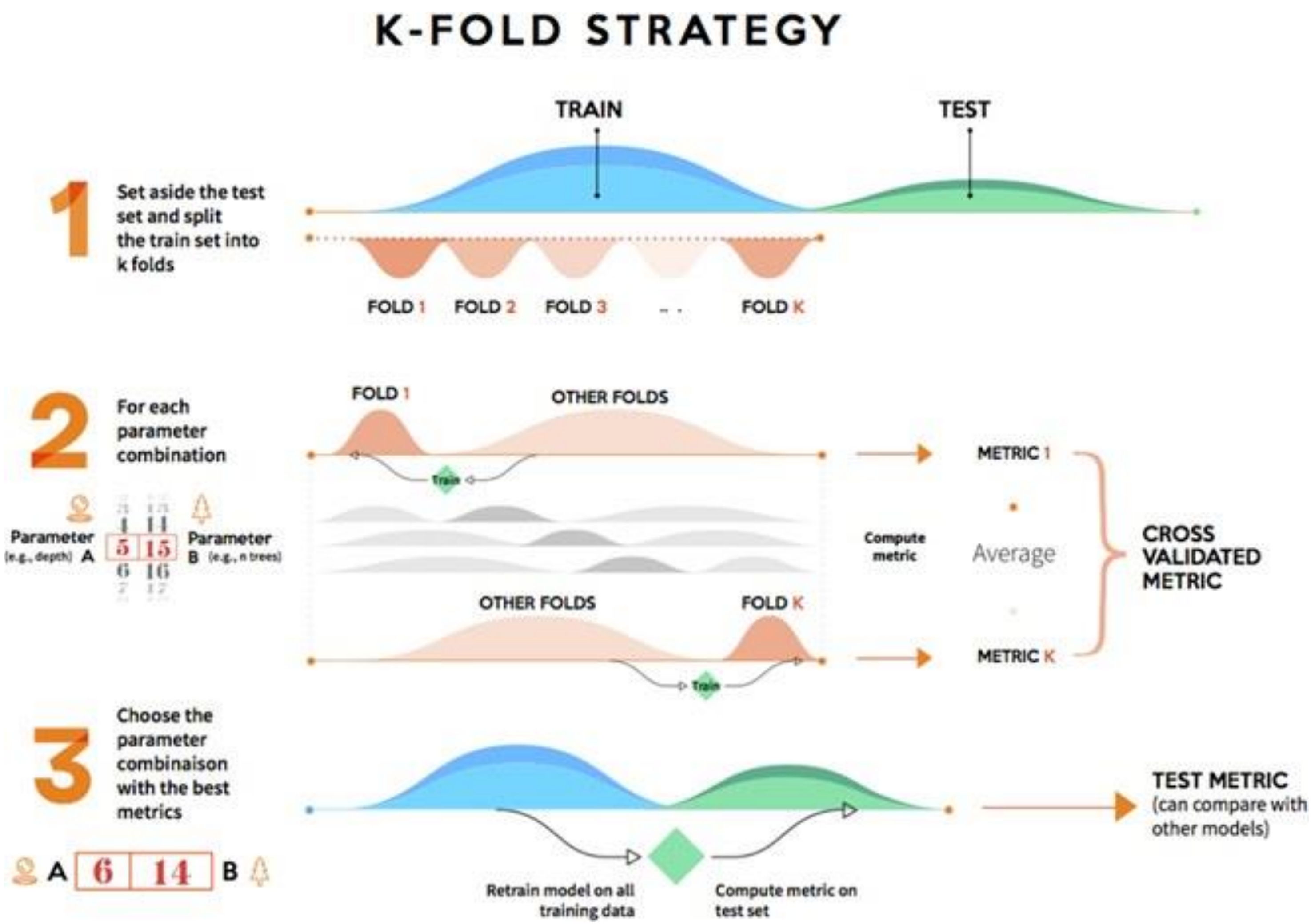
- If model being fitted is linear or polynomial regression, a simple adjustments makes a single model fit work

$$\mathrm{CV}_{(n)} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{1 - h_i}\right)^2,$$

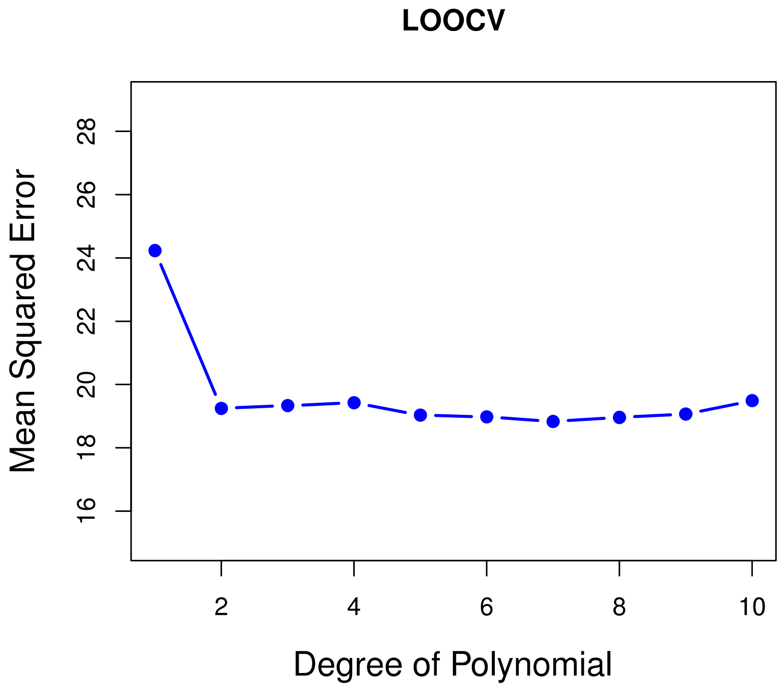
y_i is the prediction from the original least squares fit and h_i is leverage

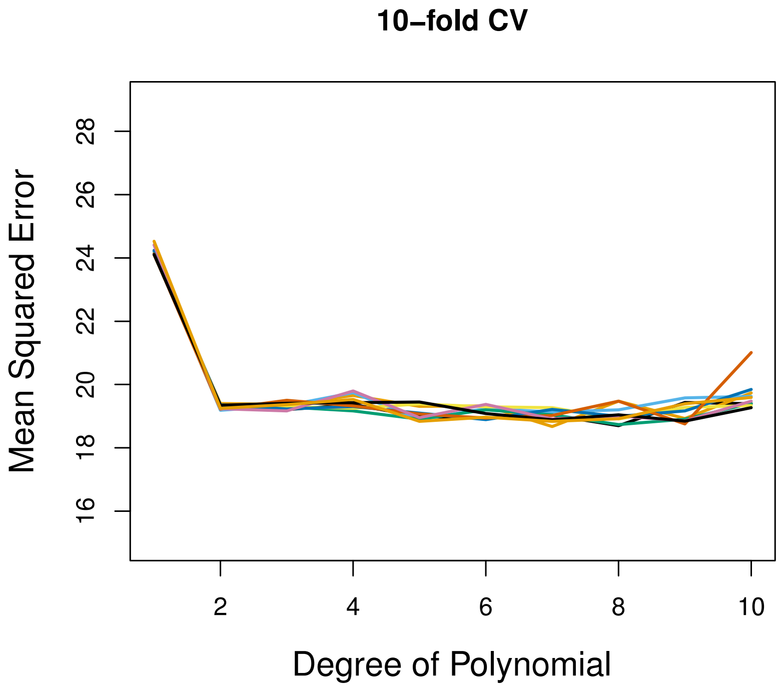
- Does not work for other models

k-Fold Cross Validation

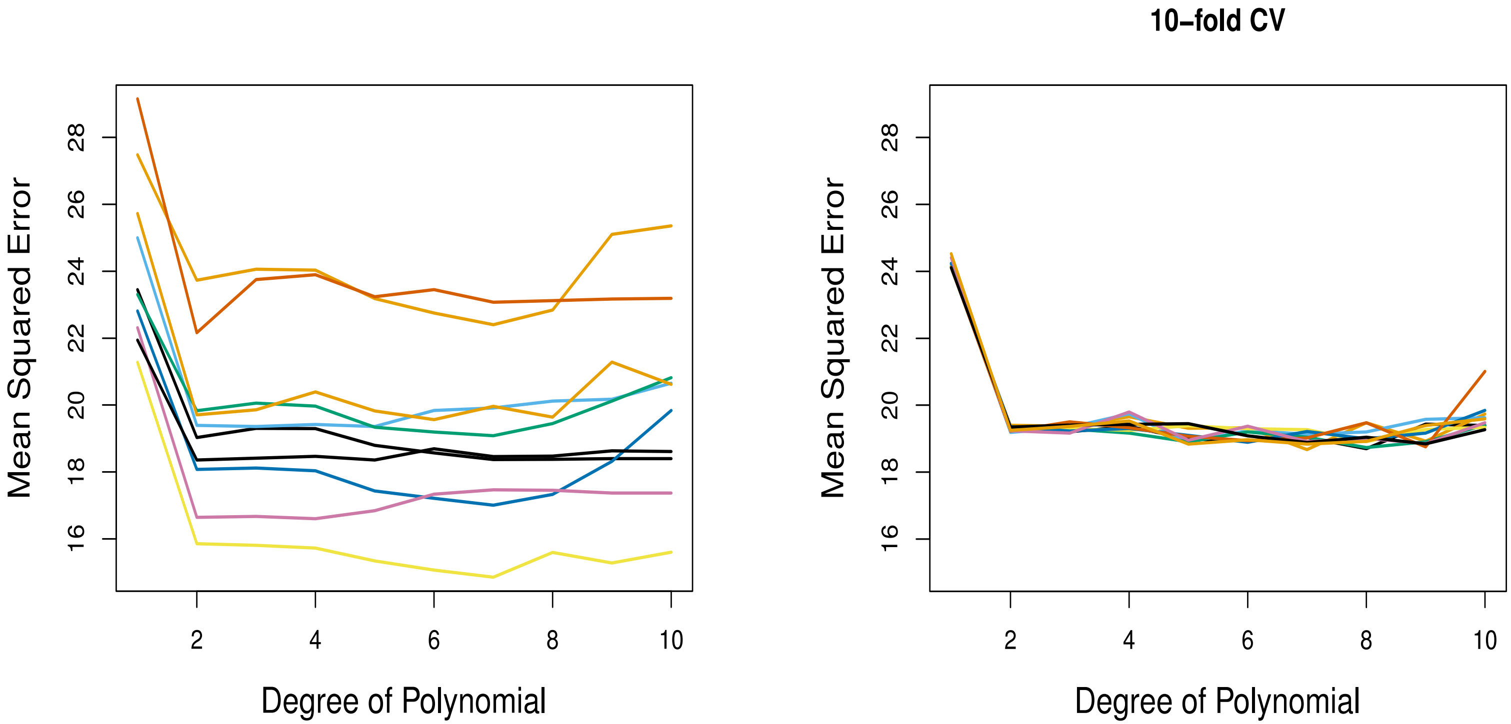


LOOCV vs. k-Fold Cross Validation





Hold Out vs. k-Fold Cross Validation



Bias-Variance Tradeoff for k-fold CV

- So which strategy is better? LOOCV or k-fold CV?
- Answer lies in Bias-Variance tradeoff
- LOOCV is less biased than k-fold CV (when $k < n$) → low bias
- LOOCV is more variable than k-fold CV (when $k < n$) → high variance
- This implies that there is a tradeoff here between bias and variance

Which One to Use?

- Sweet spot is somewhere in the middle which achieves a compromise between bias and variance
- Use k-fold cross-validation using $k = 5$ or $k = 10$
- These values have been shown empirically to yield test error rate estimates that suffer neither from excessively high bias nor from very high variance

Cross Validation for Classification

- Cross validation can also be a very useful approach in the classification setting when Y is qualitative

- CV works like described earlier; except we use misclassification error instead of MSE

- LOOCV error rate takes the form

$$\mathrm{CV}_{(n)} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{Err}_i,$$

- where $\mathrm{Err}_i = I(y_i = \hat{y}_i)$. The k-fold CV error rate and validation set error rates are defined analogously

The Bootstrap

- Powerful statistical tool that can be used to quantify the uncertainty associated with a given estimator or statistical learning method

- For example, estimate the standard errors of the coefficients from a linear regression fit

- Bootstrap is easily applied to a wide range of statistical learning methods, including some for which a measure of variability is otherwise difficult to obtain

Sampling With/Without Replacement

- Two different strategies of sampling

- **Sampling With Replacement:** Once an item is sampled, the item is placed back in the set (replaced) and then another item is drawn

- **Sampling Without Replacement:** Once an item is sampled, it is not placed back in the set and we proceed to draw the next item

Quiz

- Which sampling strategy leads to independent samples?

Example

- Say you had a population of 7 people, and you wanted to sample 2. Their names are: [John, Jack, Qiu, Tina, Hatty, Jacques, Des]

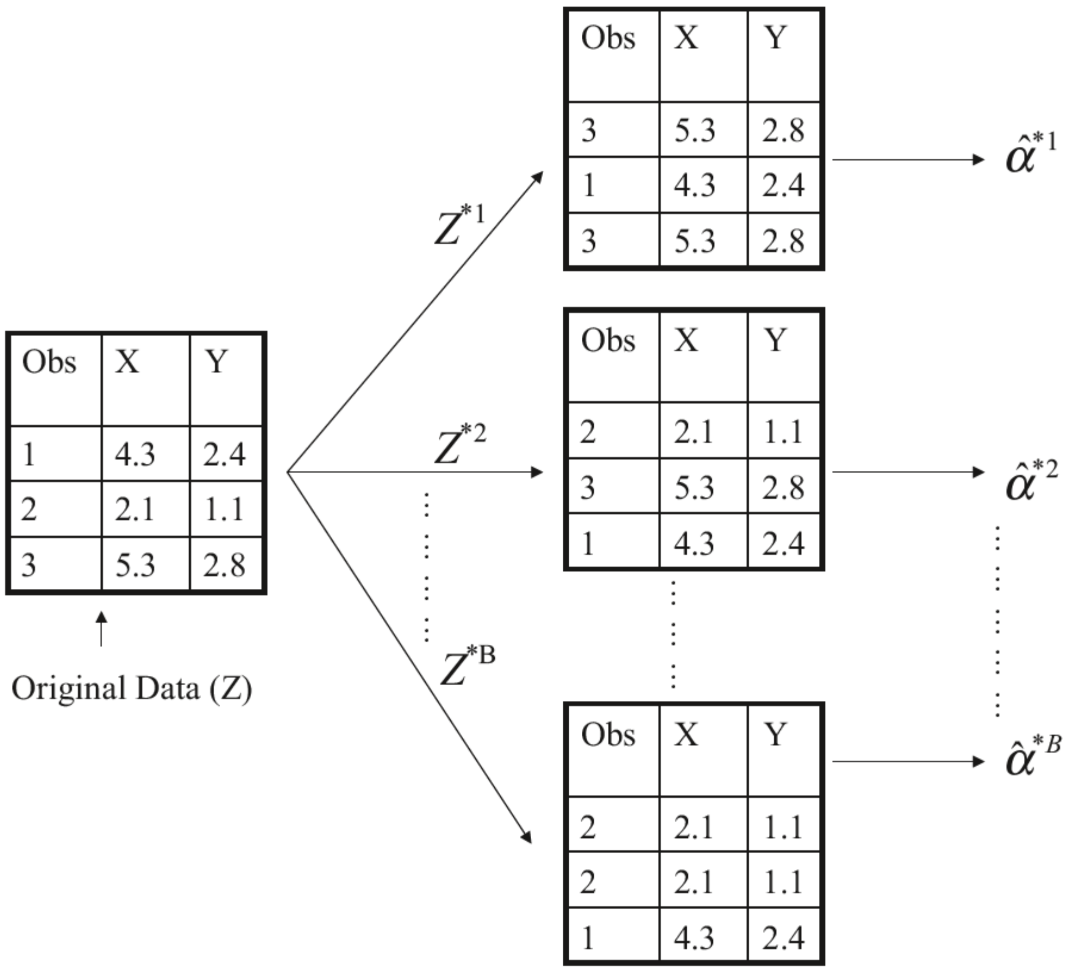
- Sampling With Replacement
  - P(John, John) = (1/7) * (1/7) = .02.
  - P(John, Jack) = (1/7) * (1/7) = .02.
  - P(John, Qui) = (1/7) * (1/7) = .02.
  - P(Jack, Qui) = (1/7) * (1/7) = .02.
  - P(Jack Tina) = (1/7) * (1/7) = .02.

- Sampling Without Replacement
  - P(John, Jack) = (1/7) * (1/6) = .024.
  - P(John, Qui) = (1/7) * (1/6) = .024.
  - P(Jack, Qui) = (1/7) * (1/6) = .024.
  - P(Jack Tina) = (1/7) * (1/6) = .024….

The Bootstrap

- Draw random samples with replacement from the training data
- Repeat the process n times to get n bootstrap data sets
- Fit the model to each of the bootstrap

The Bootstrap Approach



Other Versions of Bootstrap

- There are also other variations of the bootstrap

- Instead of simulating from the original data-set, we simulate from a distribution that was fitted using the original data-set - this is the parametric bootstrap

- We could also sample (with replacement) from the residuals of a fitted model

- There are also approaches for handling dependent data such as time-series data

When Does Bootstrap Fail?

- The bootstrap can provide faulty inference in the following situations

- Too little data - e.g. suppose there is just one data-point!

- When dealing with heavy-tailed distributions

- When the data is not IID

Resources

- Resampling Methods for Meta-Model Validation with Recommendations for Evolutionary Computation

Jupyter Notebook

- *Case Study*

ON-BRAND STATEMENT

FOR GENERAL USE

> What defines the students and faculty of the University of Washington? Above all, it’s our belief in possibility and our unshakable optimism. It’s a connection to others, both near and far. It’s a hunger that pushes us to tackle challenges and pursue progress. It’s the conviction that together we can create a world of good. And it’s our determination to Be Boundless. Join the journey at **uw.edu**.