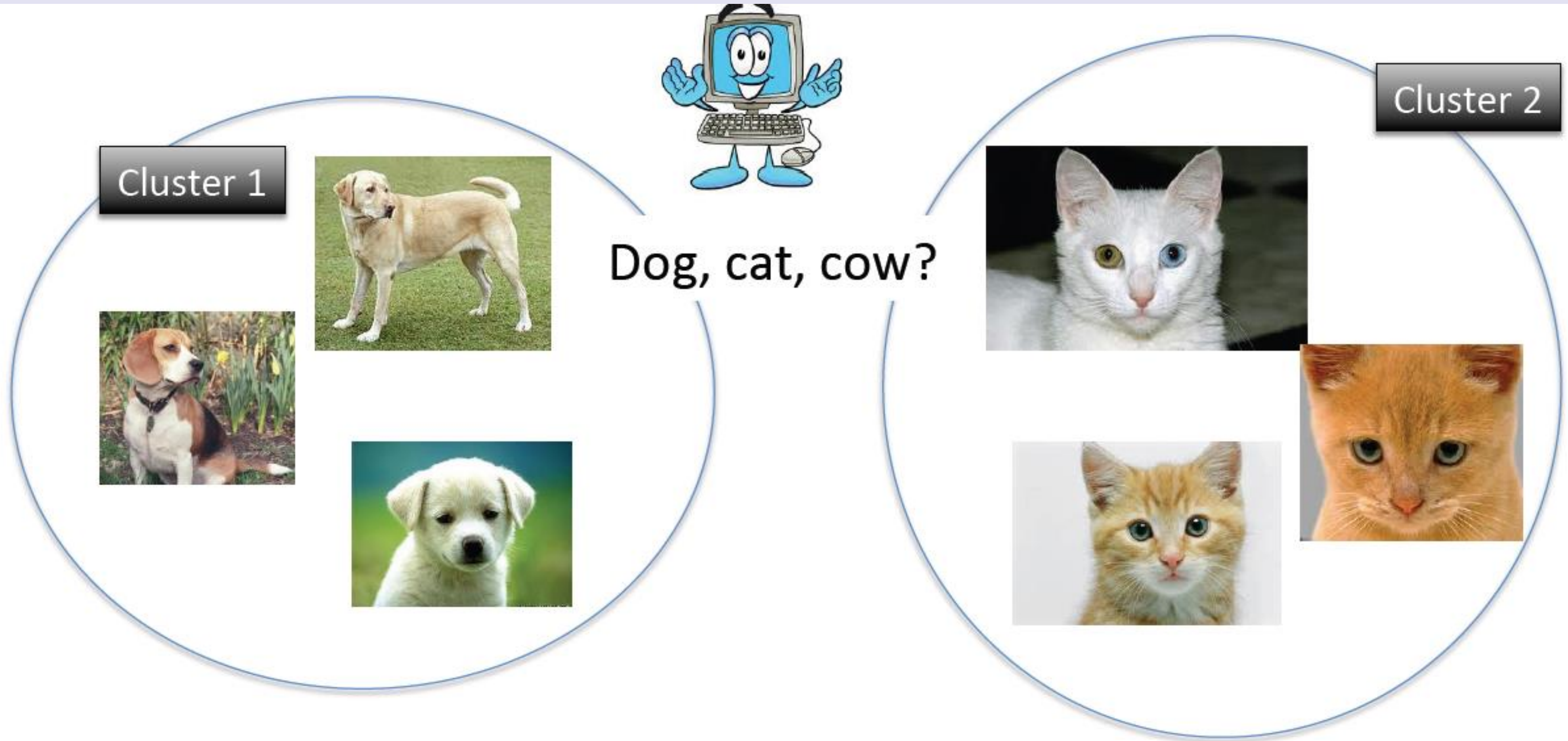
Variable selection and shrinkage methods

Course: Analytics, Machine Learning, and the Digital Economy

- **Lecturer Radjabova Dilnora**

Unsupervised Learning



Unsupervised: semantic meanings of clusters are not clear

Unsupervised Learning (cont.)

- The training data does not contain any output information at all (i.e. unlabeled data).

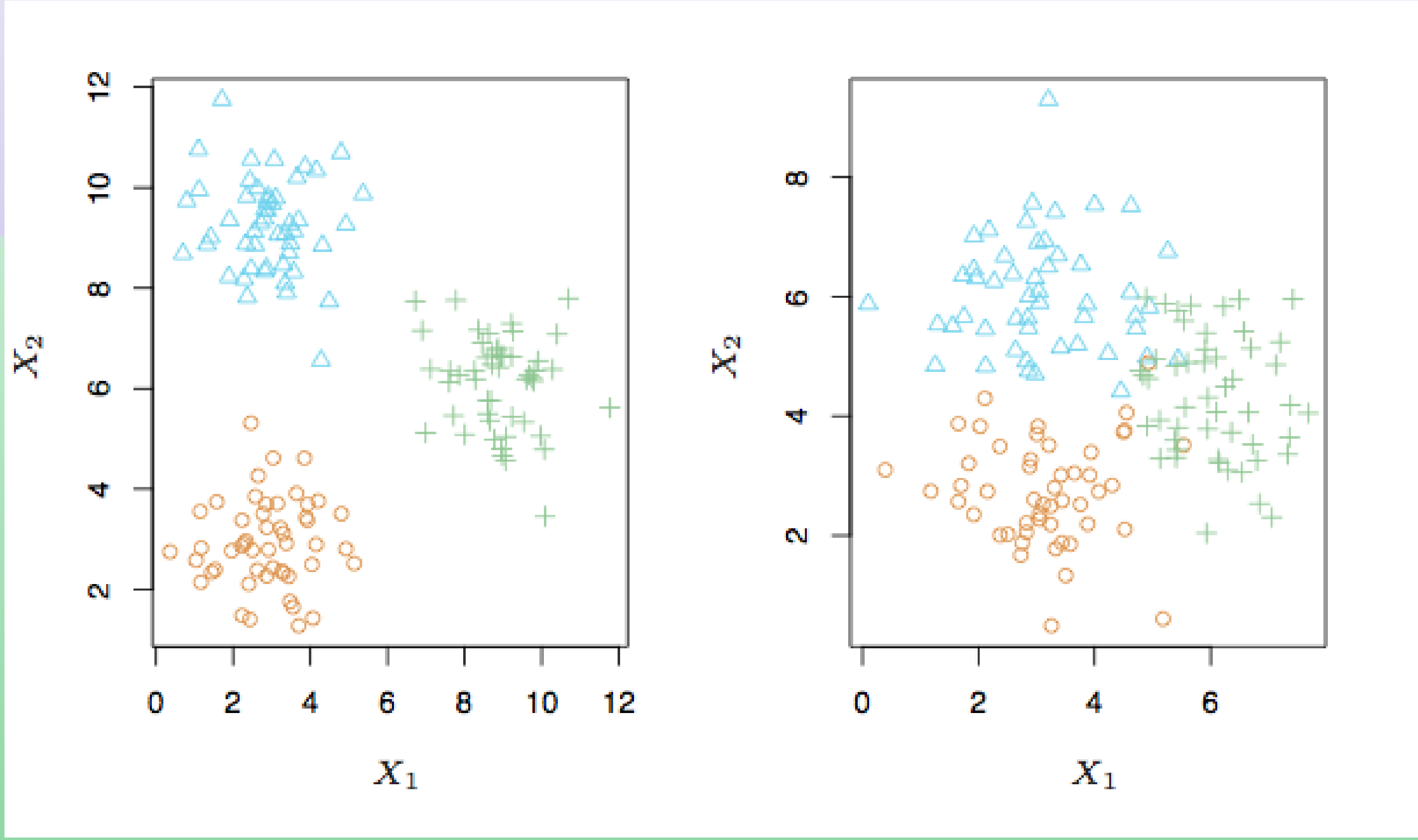
Unsupervised Learning (cont.)

- Viewed as the task of spontaneously finding patterns and structure in input data.

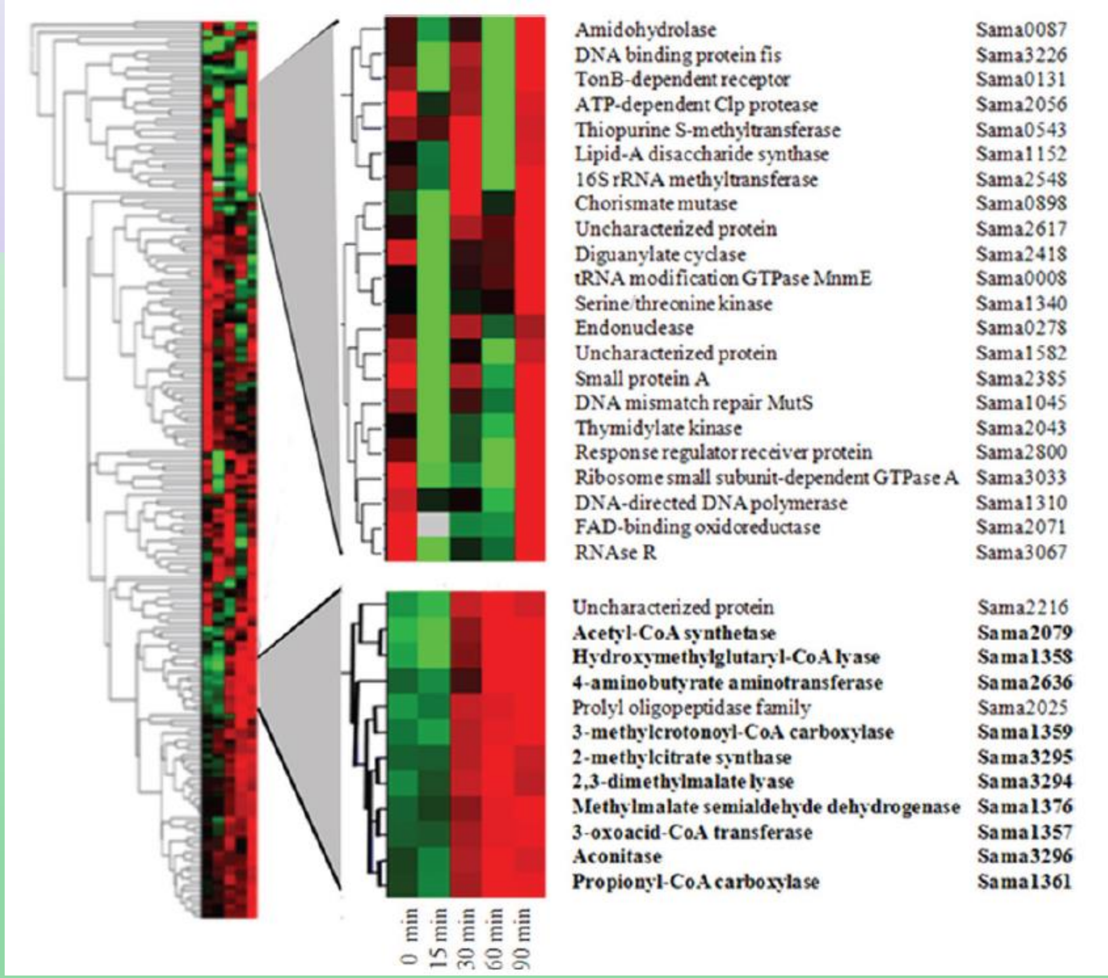
Unsupervised Learning (cont.)

- Viewed as a way to create a higher-level representation of the data and dimension reduction.

Unsupervised Learning: K-Means Clustering



Unsupervised Learning: Hierarchical Clustering



Assessing Model Accuracy

- For a given set of data, we need to decide which machine learning method produces the **best** results.

Assessing Model Accuracy

- We need some way to measure the quality of fit (i.e. how well its predictions actually match the observed data).

Assessing Model Accuracy

- In regression, we typically use mean squared error (MSE).

Assessing Model Accuracy (cont.)

Suppose we fit a model $\hat{f}(x)$ to some training data $\mathsf{Tr} = \{x_i, y_i\}_1^N$, and we wish to see how well it performs.

- We could compute the average squared prediction error over Tr:

$$\mathrm{MSE}_{\mathsf{Tr}} = \mathrm{Ave}_{i \in \mathsf{Tr}}[y_i - \hat{f}(x_i)]^2$$

This may be biased toward more overfit models.

- Instead we should, if possible, compute it using fresh *test* data $\mathsf{Te} = \{x_i, y_i\}_1^M$:

$$\mathrm{MSE}_{\mathsf{Te}} = \mathrm{Ave}_{i \in \mathsf{Te}}[y_i - \hat{f}(x_i)]^2$$

Assessing Model Accuracy (cont.)

- There is no guarantee that the method with the smallest *training MSE* will have the smallest *test MSE*.

Assessing Model Accuracy (cont.)

- There is no guarantee that the method with the smallest *training MSE* will have the smallest *test MSE*.

Training vs. Test MSEs

- In general, the more flexible a method is the lower its training MSE will be.

Training vs. Test MSEs

- In general, the more flexible a method is the lower its training MSE will be.

Training vs. Test MSEs

- However, the test MSE may in fact be higher for a more flexible method than for a simple approach like linear regression.

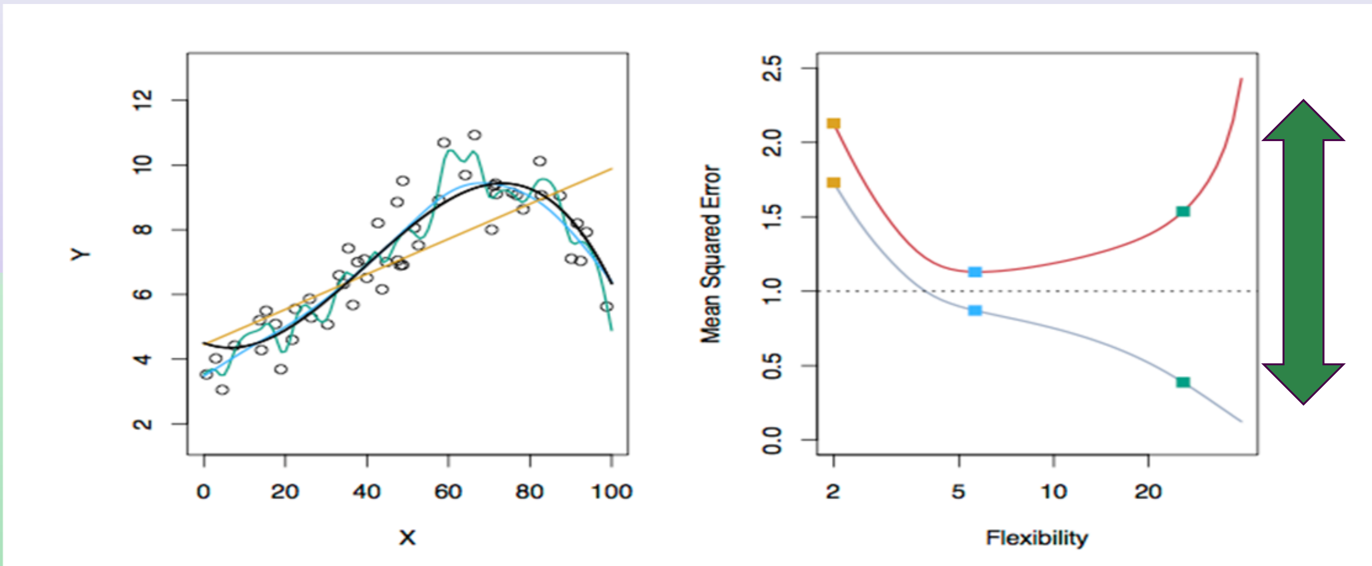
Training vs. Test MSEs (cont.)

- More flexible methods (such as splines) can generate a wider range of possible shapes to estimate f as compared to less flexible and more restrictive methods (such as linear regression).

Training vs. Test MSEs (cont.)

- The less flexible the method, the easier to interpret the model. there is a trade-off between flexibility and model interpretability.

Different Levels of Flexibility

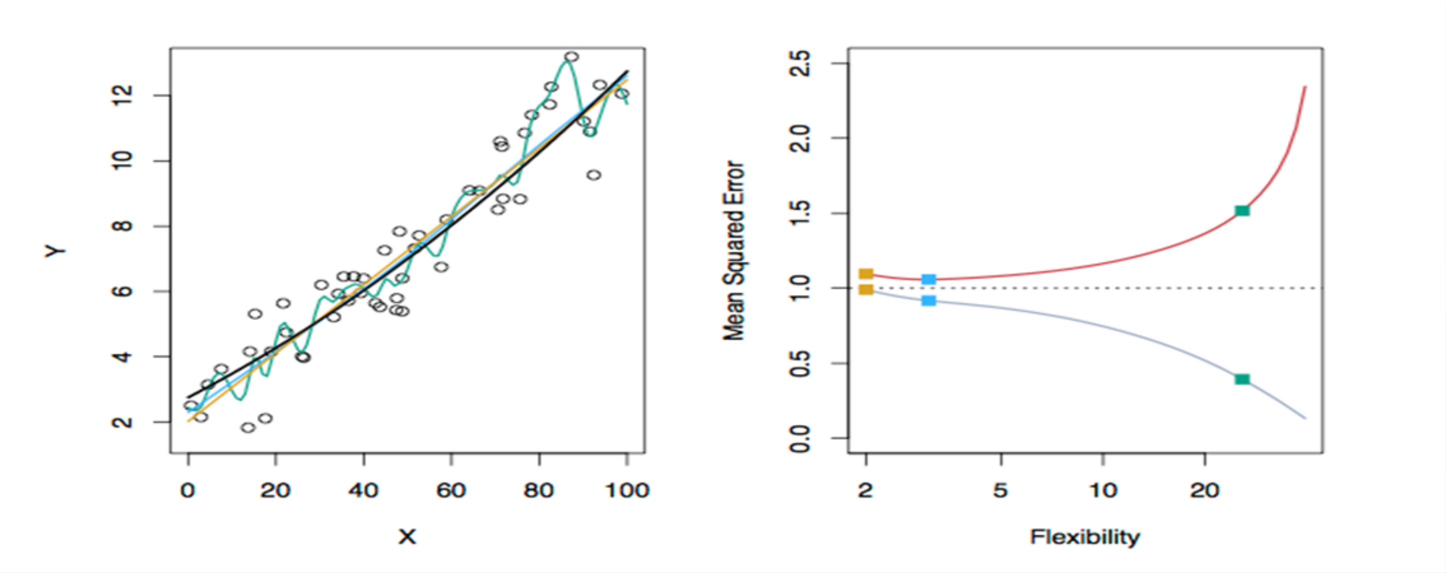


Overfitting

<u>LEFT</u>
Black: Truth
Orange: Linear estimate
Blue: Smoothing spline
Green: Smoothing spline (more flexible)

<u>RIGHT</u>
RED: Test MSE
Grey: Training MSE
Dashed: Minimum possible test MSE (irreducible error)

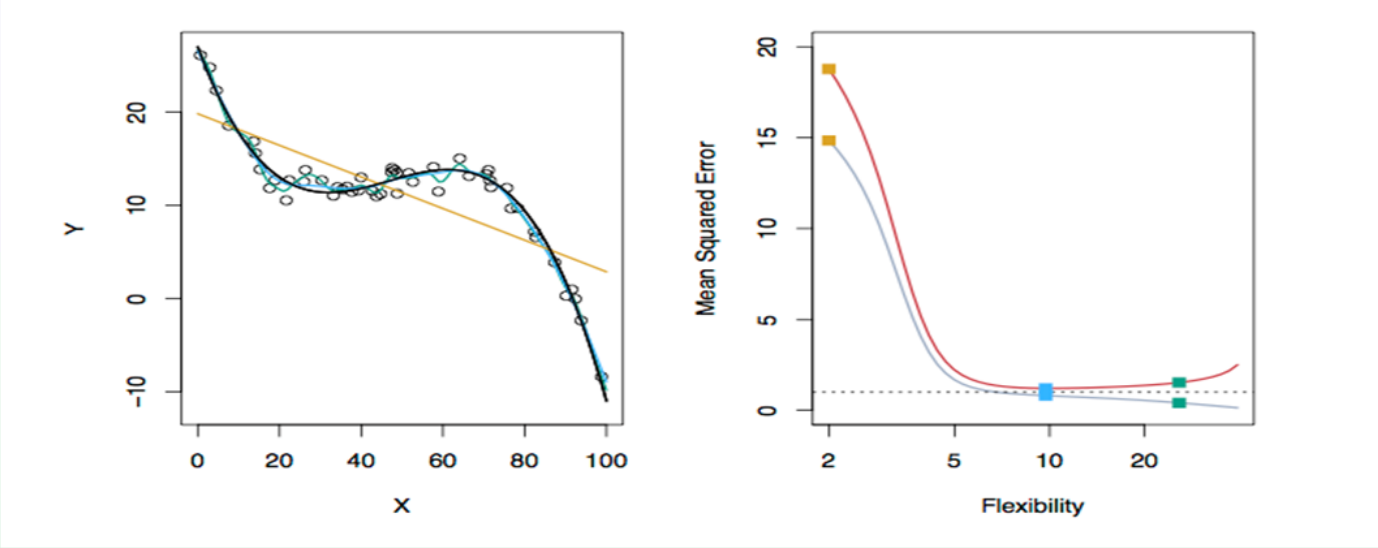
Different Levels of Flexibility (cont.)



<u>LEFT</u>
Black: Truth
Orange: Linear estimate
Blue: Smoothing spline
Green: Smoothing spline (more flexible)

<u>RIGHT</u>
RED: Test MSE
Grey: Training MSE
Dashed: Minimum possible test MSE (irreducible error)

Different Levels of Flexibility (cont.)



<u>LEFT</u>
Black: Truth
Orange: Linear estimate
Blue: Smoothing spline
Green: Smoothing spline (more flexible)

<u>RIGHT</u>
RED: Test MSE
Grey: Training MSE
Dashed: Minimum possible test MSE (irreducible error)

Bias-Variance Trade-off

- The previous graphs of test versus training MSEs illustrates a very important trade-off that governs the choice of machine learning methods.

Bias-Variance Trade-off

- There are always two competing forces that govern the choice of learning method:
  - bias and variance

Bias of Learning Methods

- Bias refers to the error that is introduced by modeling a real life problem (that is usually extremely complicated) by a much simpler problem.

Bias of Learning Methods

- Generally, the more flexible/complex a machine learning method is, the **less bias** it will generally have.

Variance of Learning Methods

- Variance refers to how much your estimate for f would change by if you had a different training data set.

- Generally, the more flexible/complex a machine learning method is the **more variance** it has.

The Trade-Off: Expected Test MSE

Suppose we have fit a model $\hat{f}(x)$ to some training data Tr, and let (x_0, y_0) be a test observation drawn from the population. If the true model is $Y = f(X) + \epsilon$ (with $f(x) = E(Y|X = x)$), then

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \text{Var}(\hat{f}(x_0)) + [\text{Bias}(\hat{f}(x_0))]^2 + \text{Var}(\epsilon).$$

The expectation averages over the variability of y_0 as well as the variability in Tr. Note that $\text{Bias}(\hat{f}(x_0))] = E[\hat{f}(x_0)] - f(x_0)$.

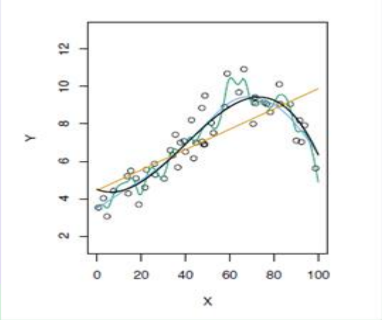
Typically as the *flexibility* of $\hat{f}$ increases, its variance increases, and its bias decreases. So choosing the flexibility based on average test error amounts to a *bias-variance trade-off.*

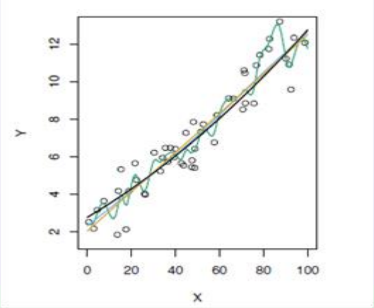
Test MSE, Bias and Variance

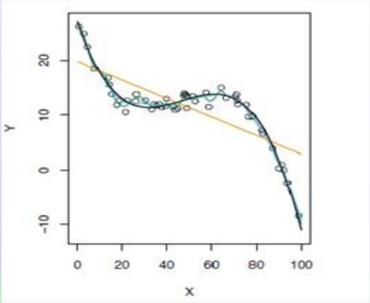
- Thus, in order to minimize the expected test MSE, we must select a machine learning method that simultaneously achieves *low variance* and *low bias*.

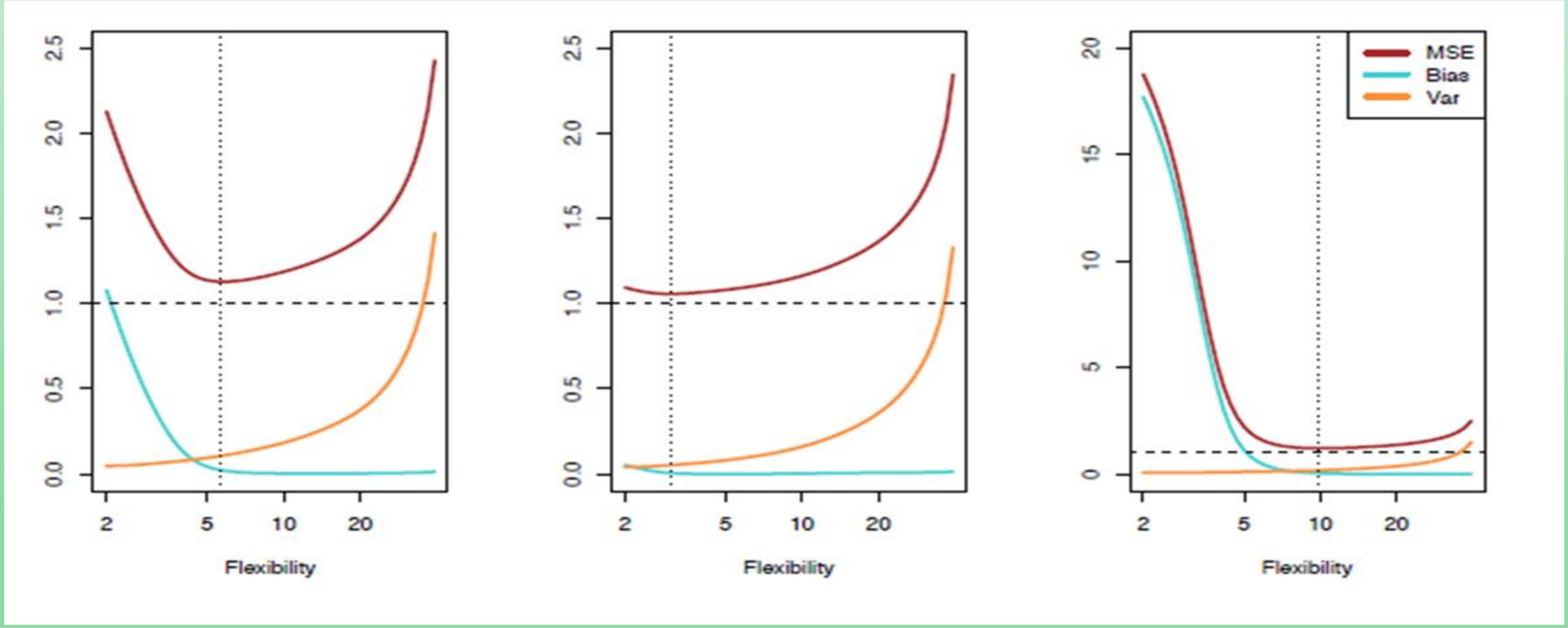
- Note that the expected test MSE can never lie below the irreducible error - Var(ε).

Test MSE, Bias and Variance (cont.)









Reference

- 1. Shalev-Shwartz and Ben-David. Understanding Machine Learning: From Theory to Algorithms (Cambridge University Press, 2014)
- 2. Daum´e. A Course in Machine Learning.
- 3. The Art of Statistics: How to Learn from Data by David Shpigelter
- 4. Learning From Data – January 1, 2012 by Yaser S. Abu-Mostafa (Author), Malik Magdon-Ismail (Author), Hsuan-Tien Lin (Author)
- 5. Statistics: The Art and Science of Learning from Data by Alan Agresti
- 6. Learning From Data: An Introduction To Statistical Reasoning by M.Glenber.
- 7. Statistics: Learning from Data (with JMP Printed Access Card) by Rocky Pek
- 8. The Elements of Statistical Learning by Gerim Garold
- 9. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 2nd Edition
- by Aurélien Géron (Author)