Assessing Model Accuracy

Prof. Asim Tewari
IIT Bombay

Measuring the Quality of Fit

Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2$$

We are interested in the accuracy of the predictions that we obtain when we apply our method to previously unseen data (Test MSE) and not on the trained data (Training MSE).

ANN

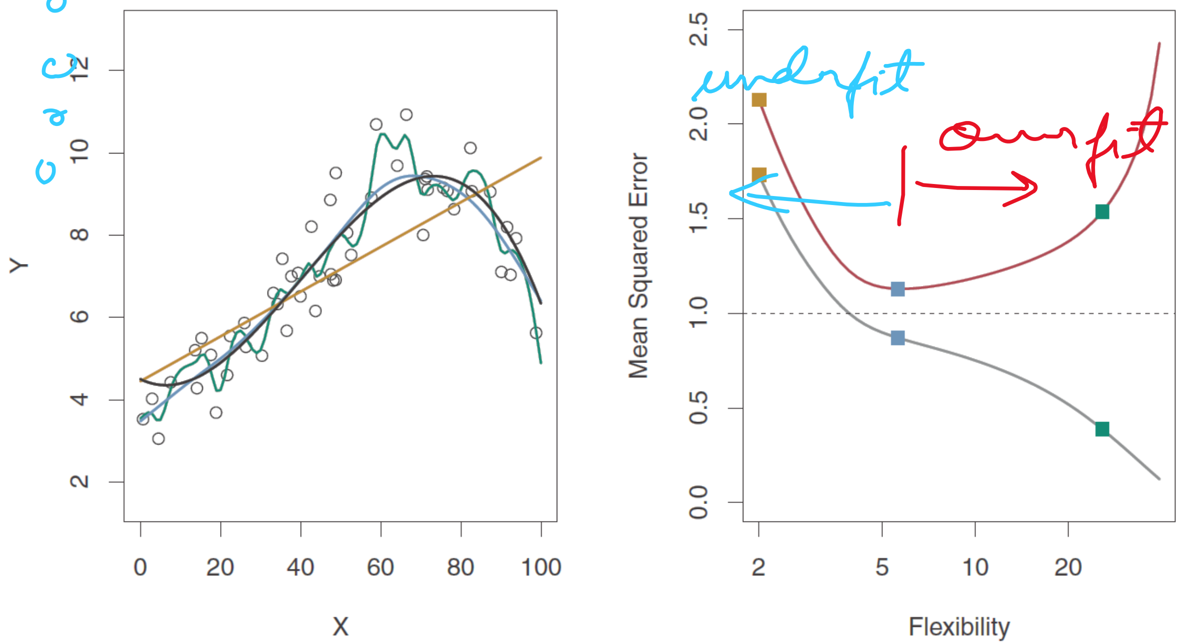
x_1

x_2

x_3

x_p

Overfitting



- Test MSE

-Training MSE

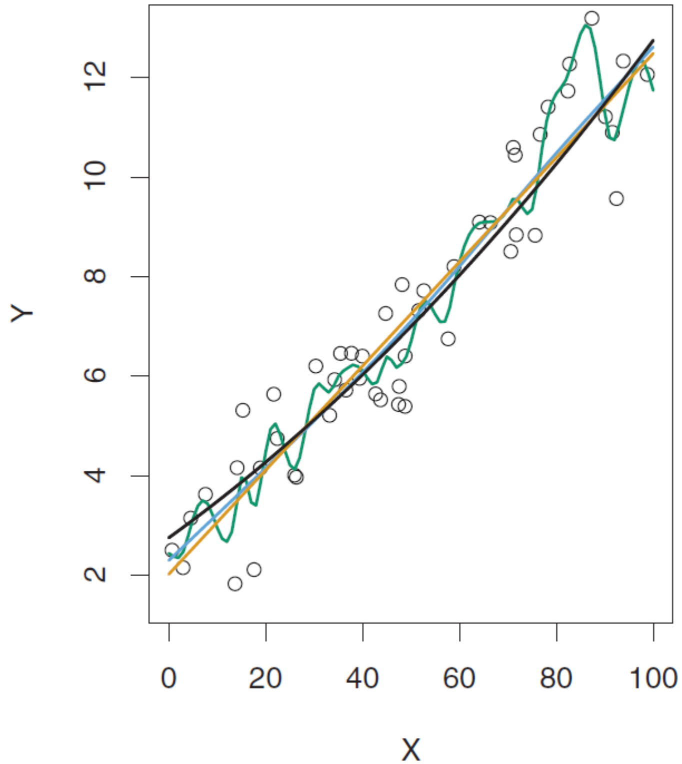
Left: Data simulated from f, shown in black. Three estimates of f are shown: the linear regression line (orange curve), and two smoothing spline fits (blue and green curves). Right: Training MSE (grey curve), test MSE (red curve), and minimum possible test MSE over all methods (dashed line). Squares represent the training and test MSEs for the three fits shown in the left-hand panel.

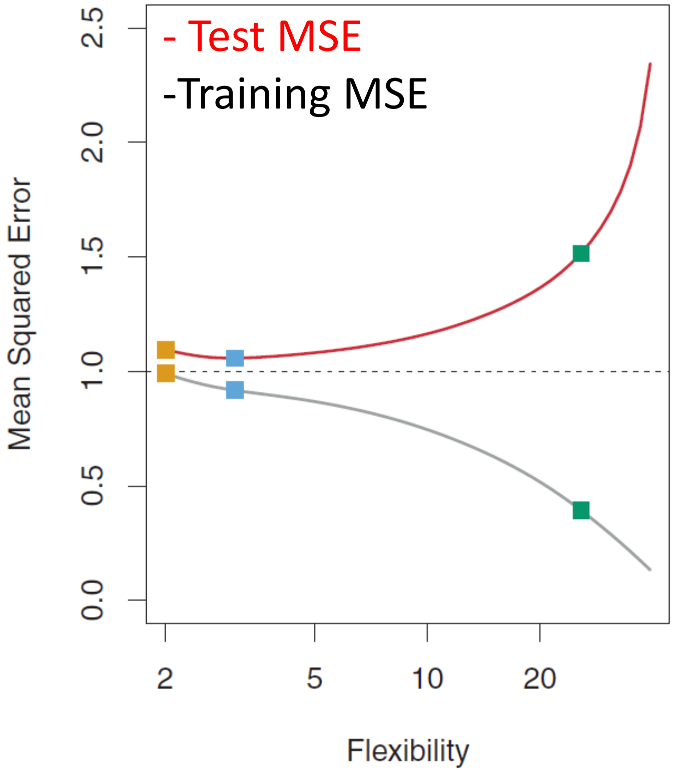
When a given method yields a small training MSE but a large test MSE, we are said to be *overfitting* the data (The U-shape in test MSE)

Overfitting

- When a given method yields a **small training MSE** but a **large test MSE**, we are said to be overfitting the data.
- Statistical learning procedure may be picking up some patterns that are just caused by random chance rather than by true properties of the unknown function f.
- Regardless of whether or not overfitting has occurred, we almost always expect the training MSE to be smaller than the test MSE because most statistical learning methods either directly or indirectly seek to minimize the training MSE.
- Overfitting refers specifically to the case in which a less flexible model would have yielded a smaller test MSE.

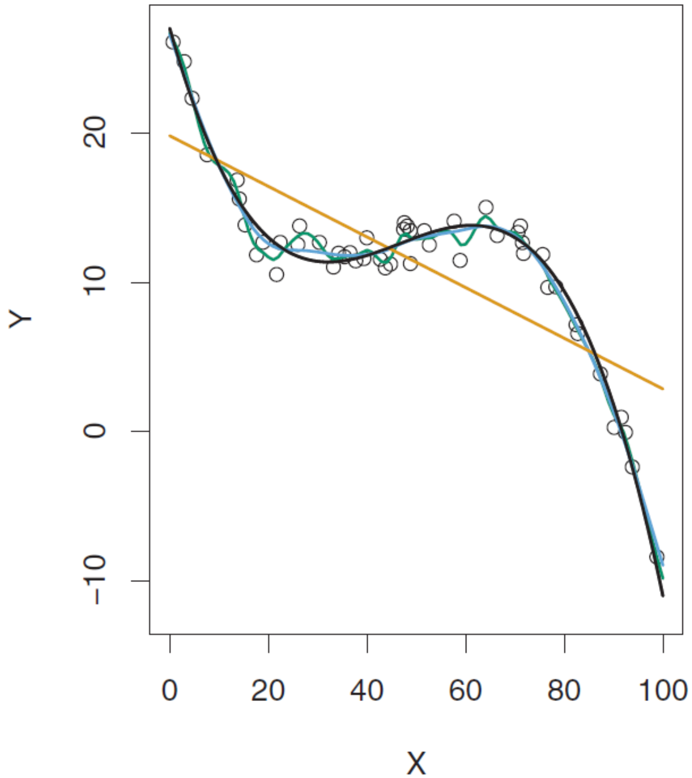
Overfitting examples

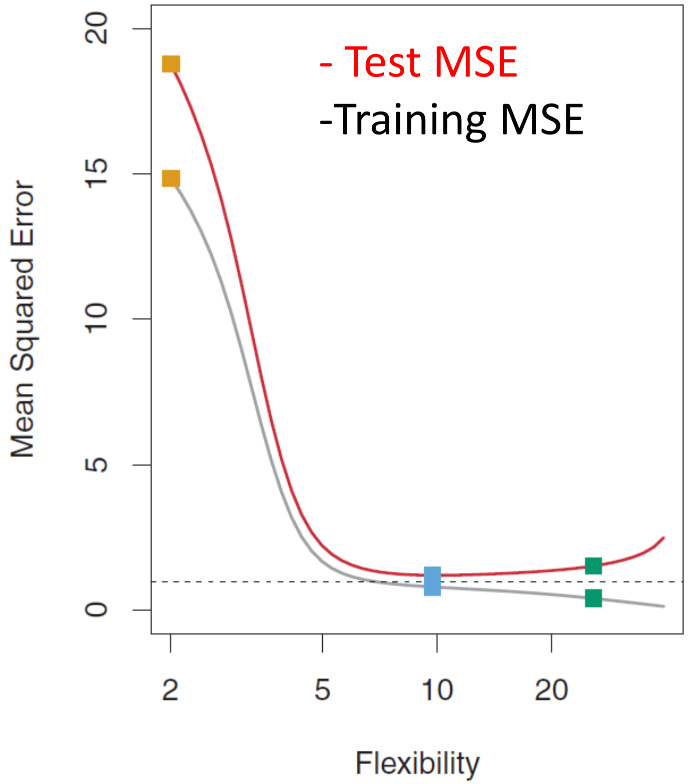




A different true f that is much closer to linear. In this setting, linear regression provides a very good fit to the data.

Overfitting examples



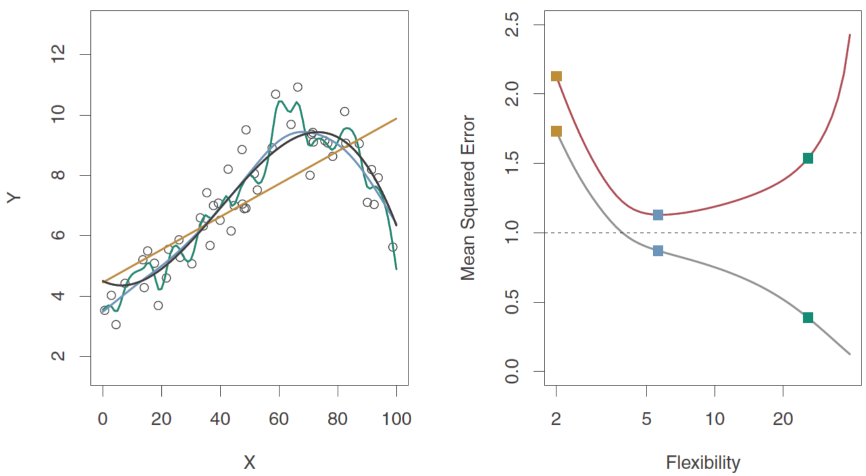


A different f that is far from linear. In this setting, linear regression provides a very poor fit to the data.

Overfitting examples

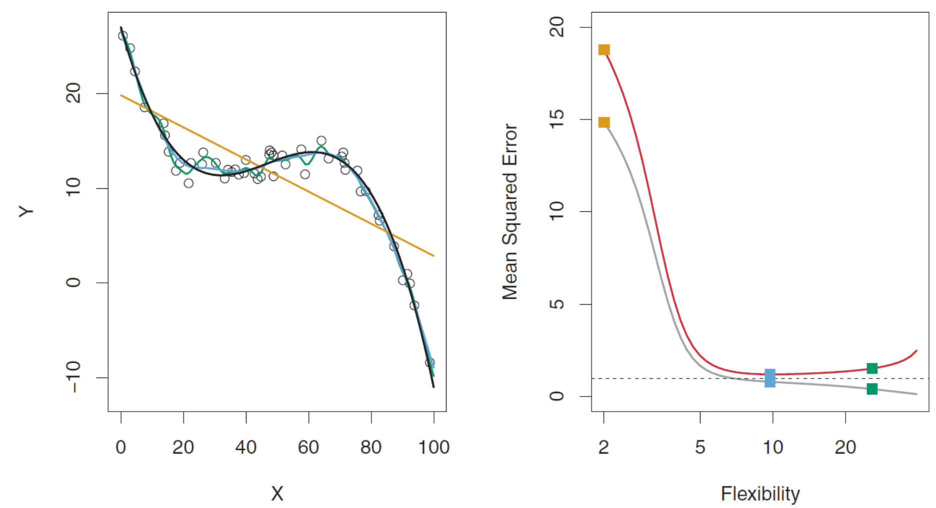
- Test MSE

-Training MSE



- Test MSE

-Training MSE



A different f that is far from linear. In this setting, linear regression provides a very poor fit to the data.

The Bias-Variance Trade-Off

The U-shape observed in the test MSE curves is a result of two competing properties of statistical learning methods.

It can be shown that that the expected test MSE, for a given value *x*0 is given by:

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

In order to minimize the expected test error, we need to select a statistical learning method that simultaneously achieves low variance and low bias.

Training and Testing

The Bias-Variance Trade-Off

The U-shape observed in the test MSE curves is a result of two competing properties of statistical learning methods.

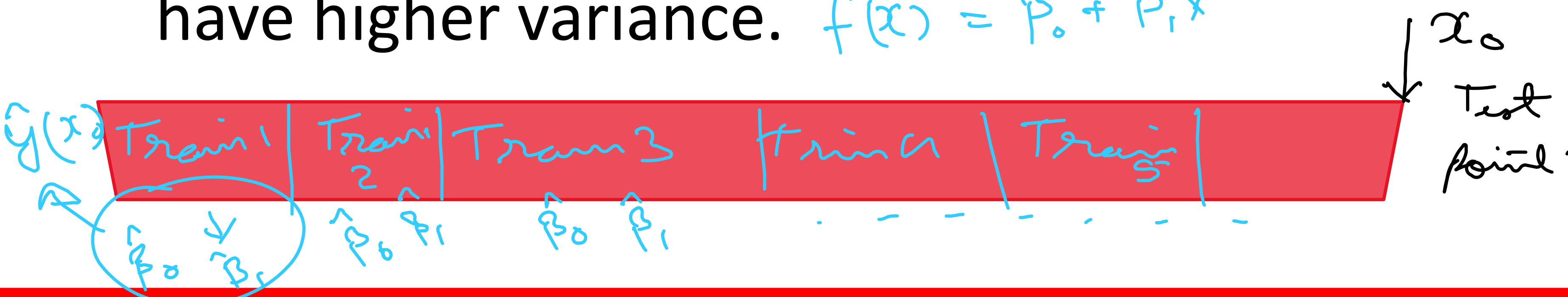
It can be shown that that the expected test MSE, for a given value *x*0 is given by:

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

In order to minimize the expected test error, we need to select a statistical learning method that simultaneously achieves low variance and low bias.

The Bias-Variance Trade-Off

- Variance refers to the amount by which $\hat{f}(x_0)$would change if we estimated it using a different training data set.
- Ideally the estimate for f should not vary too much between training sets.
- In general, more flexible statistical methods have higher variance. $\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}_1 X$

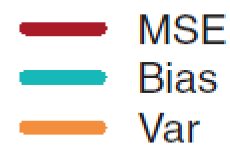


The Bias-Variance Trade-Off

- **Bias refers** to the error that is introduced by approximating a real-life problem, which may be extremely complicated, by a much simpler model.
- Bias will (on an average) either over predict or under predict the results.
- Generally, more flexible methods result in less bias.

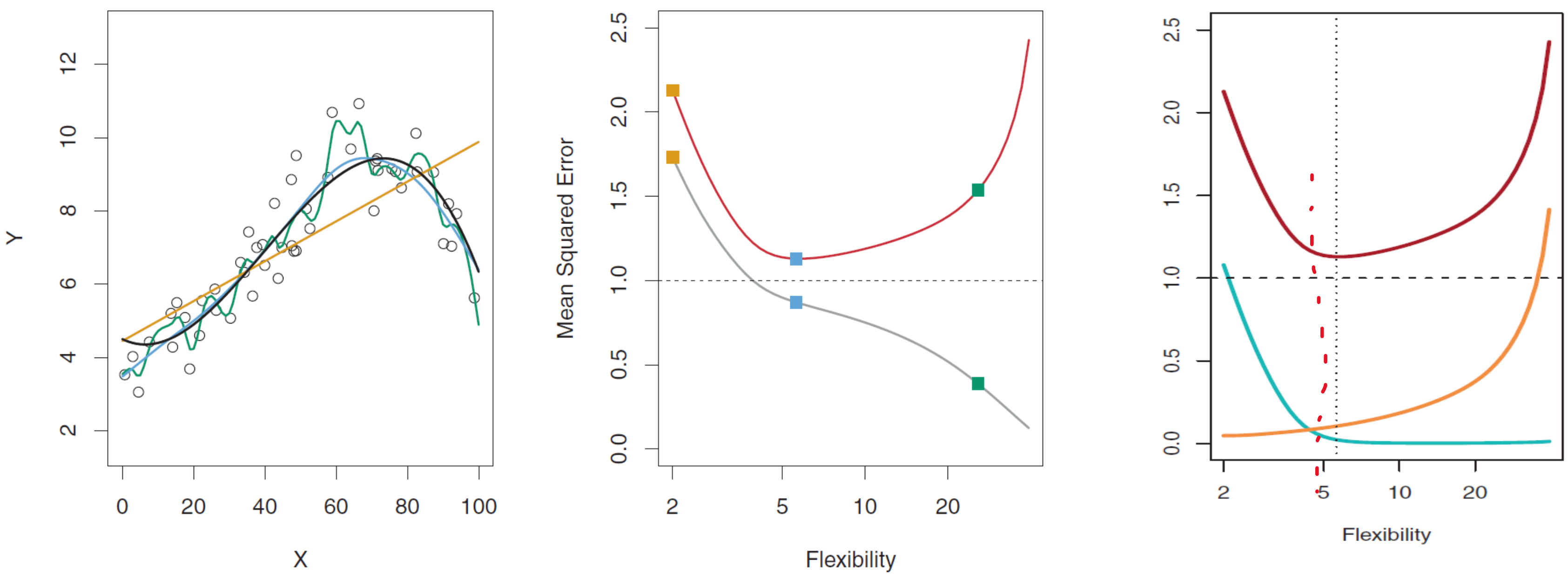
The Bias-Variance Trade-Off

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



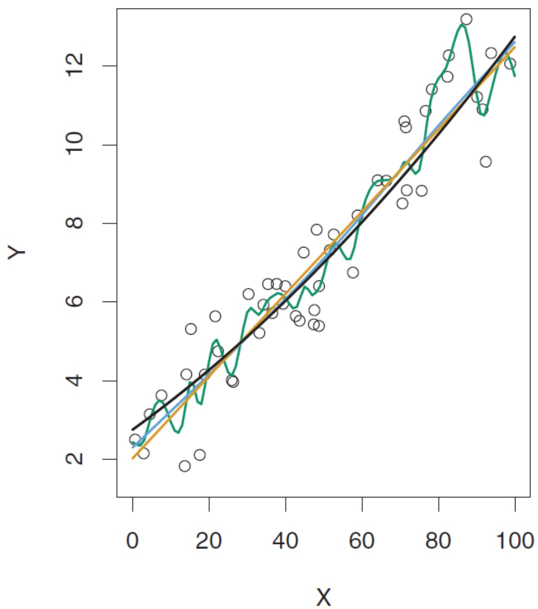
- Test MSE

-Training MSE



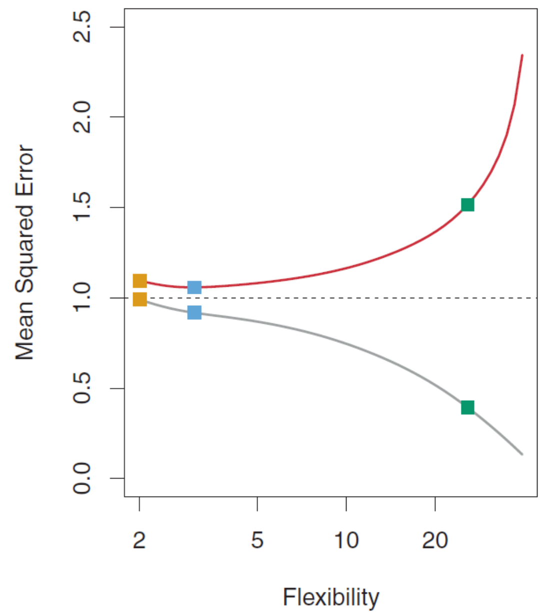
The vertical dotted line indicates the flexibility level corresponding to the smallest test MSE.

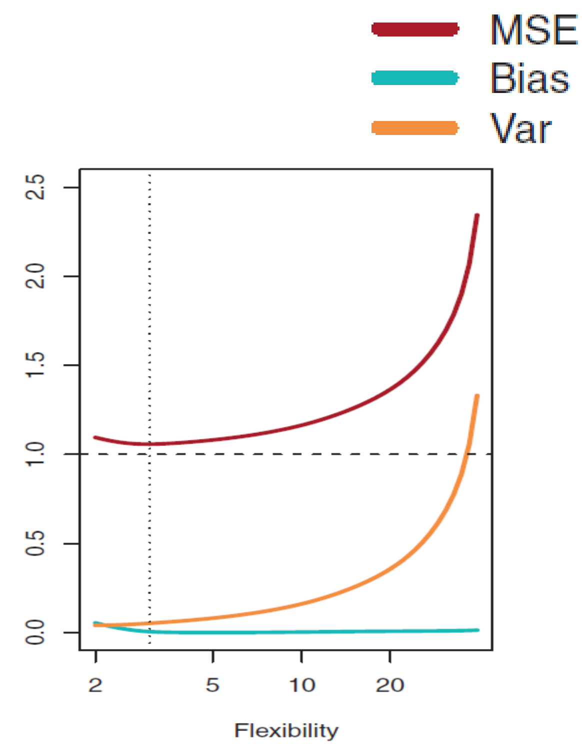
The Bias-Variance Trade-Off



- Test MSE

-Training MSE





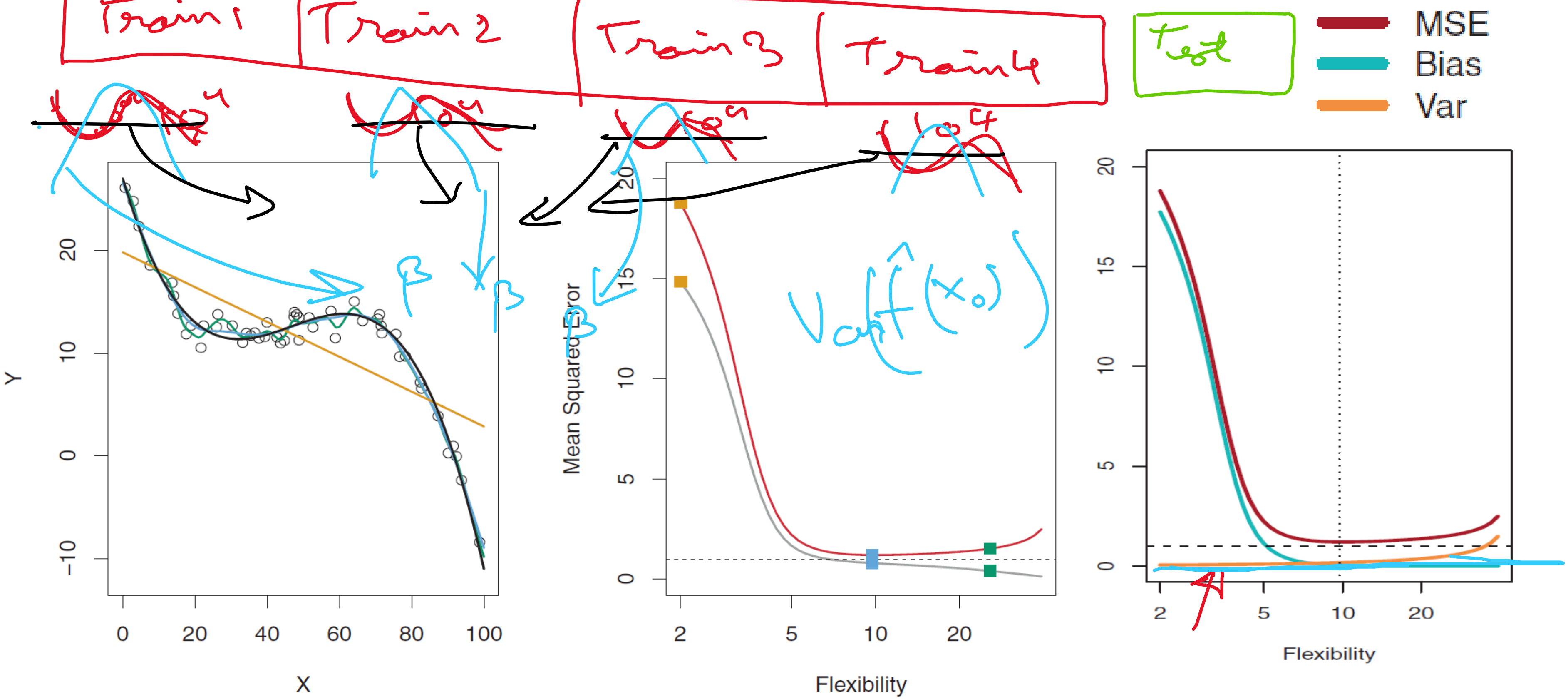
The vertical dotted line indicates the flexibility level corresponding to the smallest test MSE.

The Bias-Variance Trade-Off

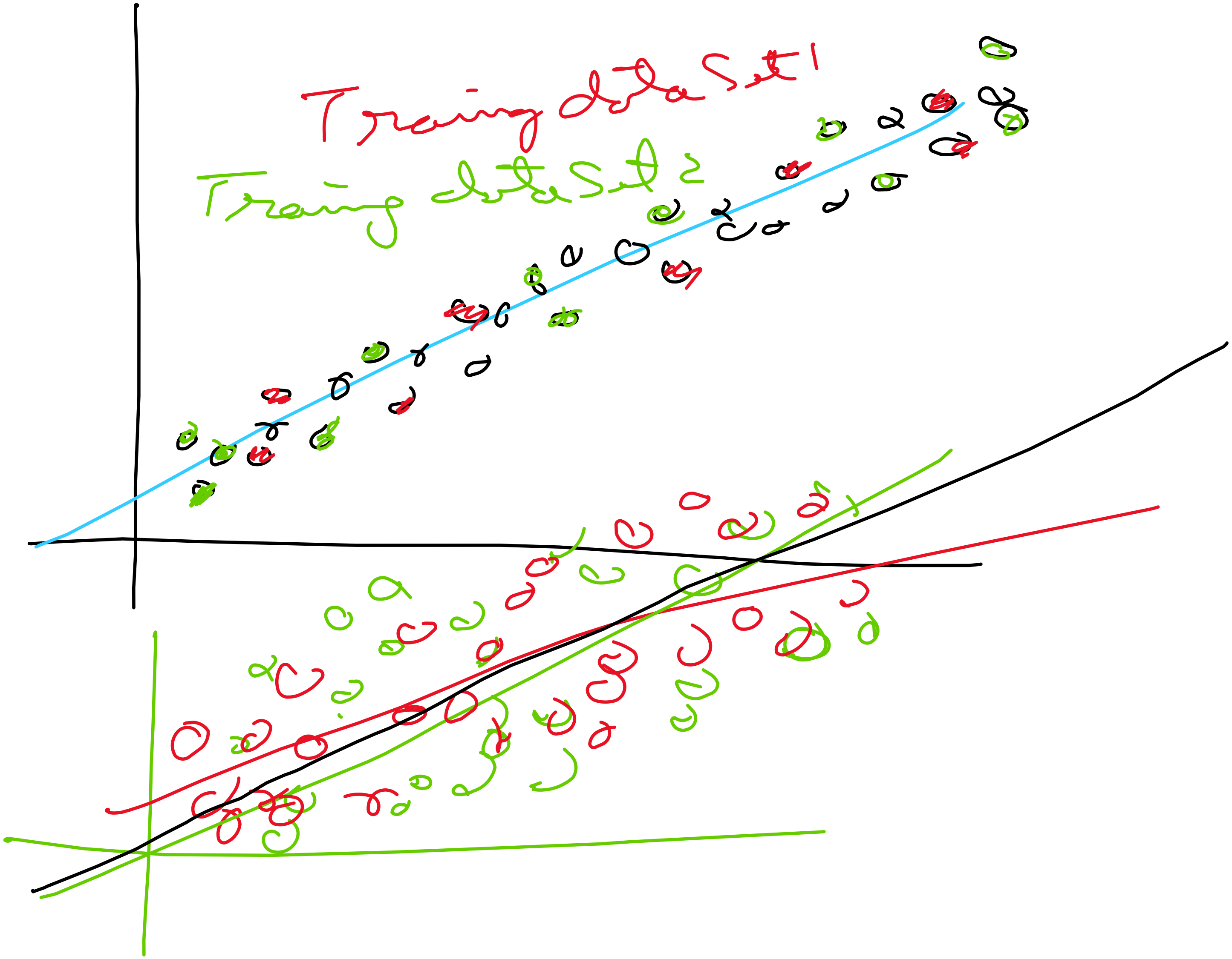
Random set

Train 1 Train 2 Train 3 Train 4

Test



The vertical dotted line indicates the flexibility level corresponding to the smallest test MSE.


Training data Set 1
Training data Set 2