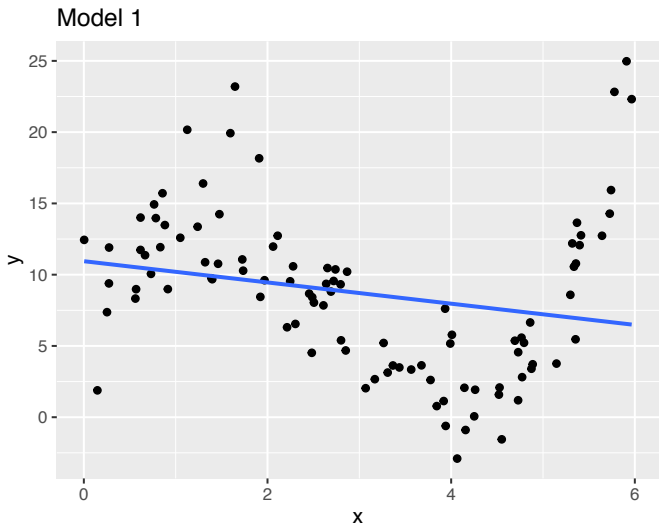
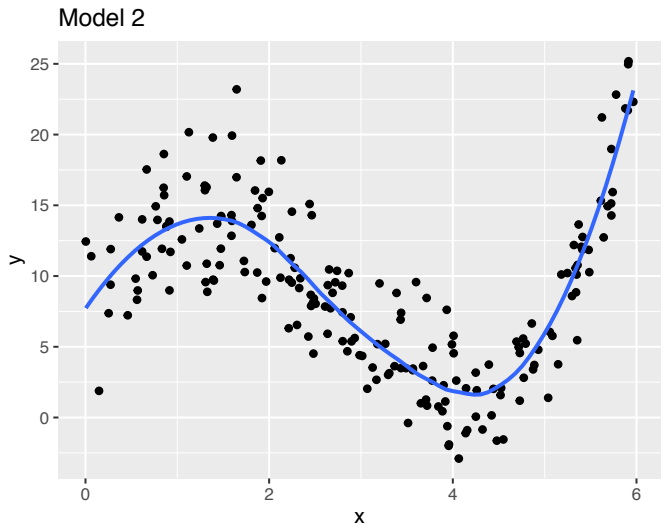
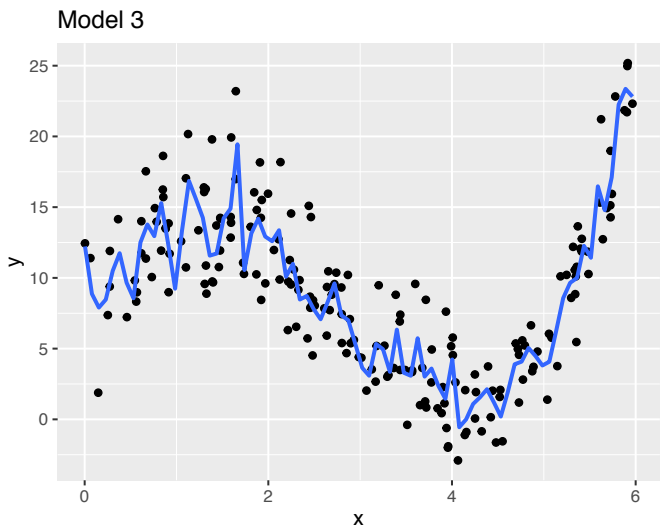
Lecture 7:
Overfitting and Bias-Variance Tradeoff
STAT 452, Spring 2021

Which model do you think is the "best fit" to the data, and will yield the most accurate predictions for new, future values of the response variable?





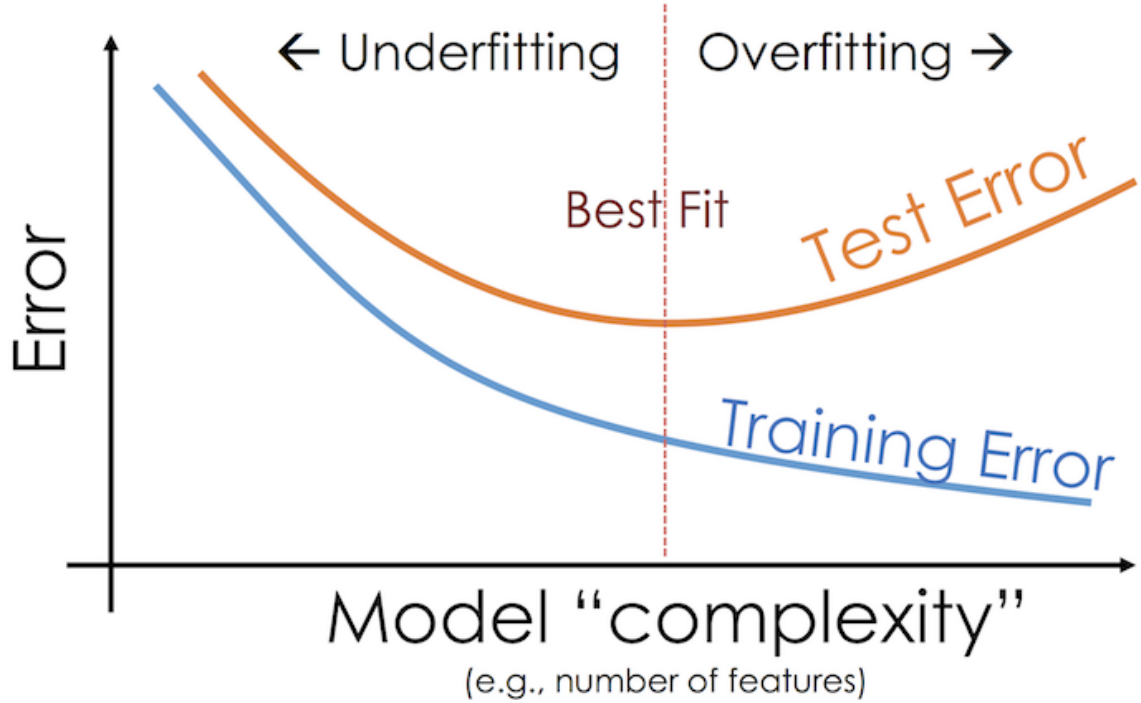


Overfitting

- Models that overfit the data will yield a small MSE on the training set but a large MSE on the test set.
- This is happening because the model is picking up patterns in training data that are caused by random chance (noise).

Overfitting

Generally, as model flexibility (or complexity) increases, the training MSE will decrease, but the test MSE may start to increase.



Overfitting

- Going back to the example, the data was split into a training set with 100 points, and a test set with 100 points. Note that the data set was simulated.
- The table below shows the results for the MSE on the training data, and the MSE on the test data.
- For model 3, which was overfitting, the test MSE is much larger than the training MSE. Model 2 has the lowest test MSE.

	Training MSE	Test MSE
Model 1	30.5	35.9
Model 2	8.3	8.5
Model 3	1.2	30.8

Bias-Variance Trade-Off

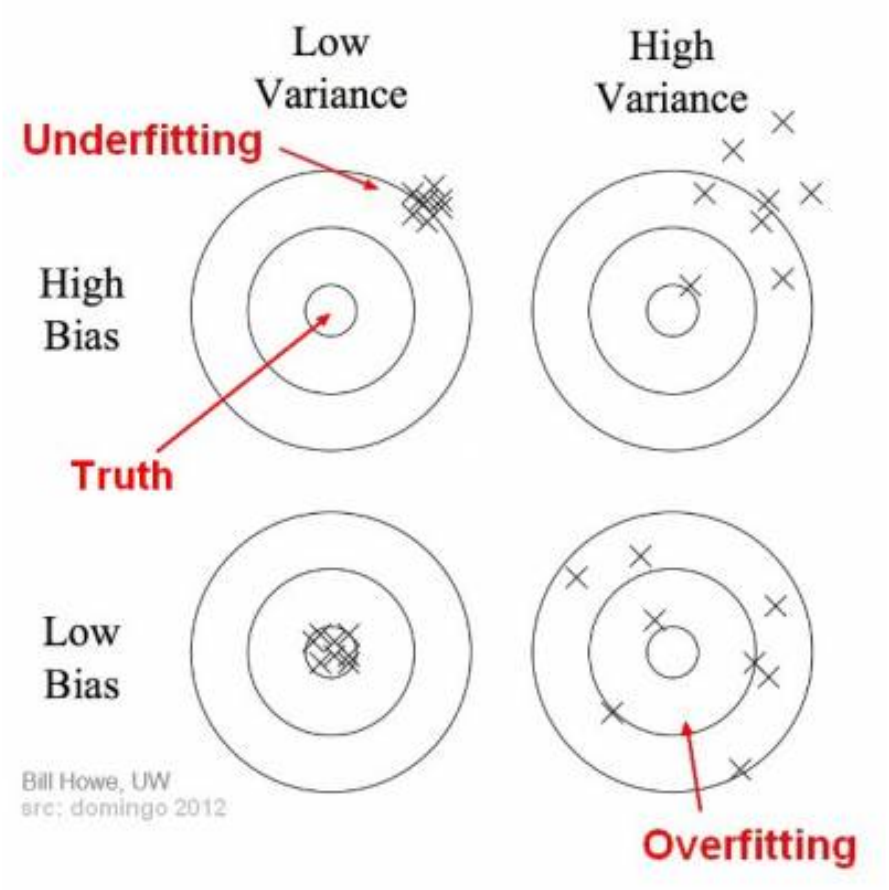
- *Variance* refers to the amount by which the estimate of the model would change if using a different training data set. The more flexible or complex the model, the higher the variance.

- *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-Variance Trade-Off

- A model that overfits the data is said to have high variance and low bias.
- A model the underfits the data is said to have low variance and high bias.
- Ideally, we want to find a model that minimizes both the bias and variance, but the two goals are usually mutually exclusive.

Bias-Variance Trade-Off



References

Chapter 2, pp. 29–36, from *An Introduction to Statistical Learning*

`https://en.wikipedia.org/wiki/Overfitting`

`https://en.wikipedia.org/wiki/Bias%E2%80%93variance_tradeoff`