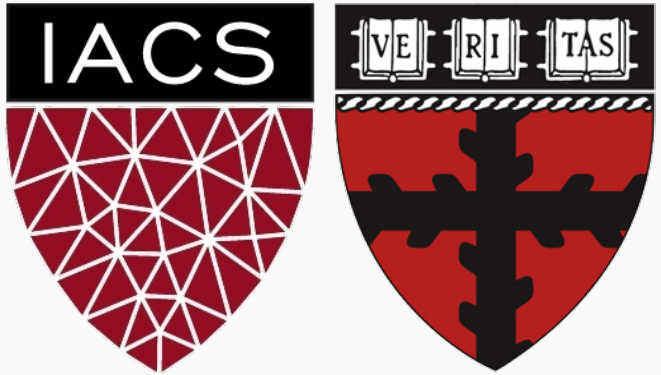
Regularization error, bias vs variance

CS109A Introduction to Data Science

Pavlos Protopapas, Kevin Rader and Chris Tanner

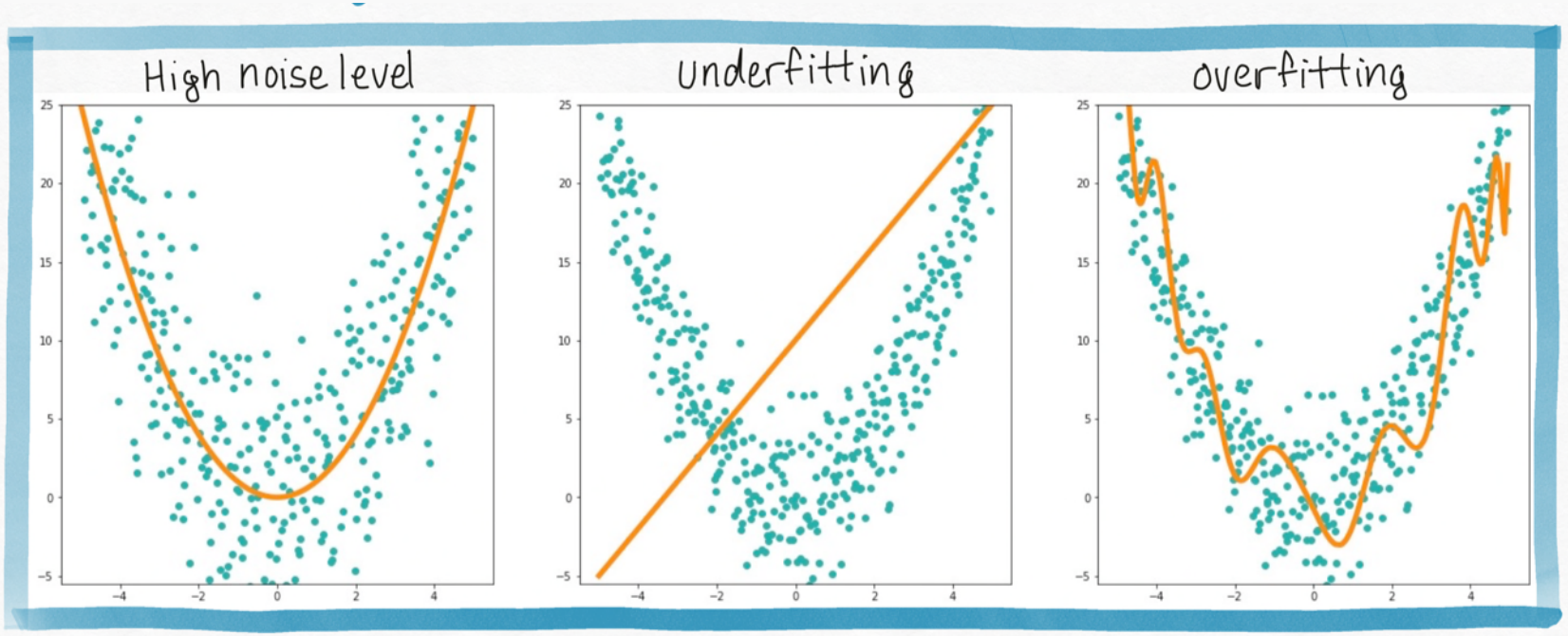


Test Error and Generalization

We know to evaluate models on both train and test data because models can do well on training data but do poorly on new data.

When models do well on new data is called generalization.

There are at least three ways a model can have a high test error.



Irreducible and Reducible Errors

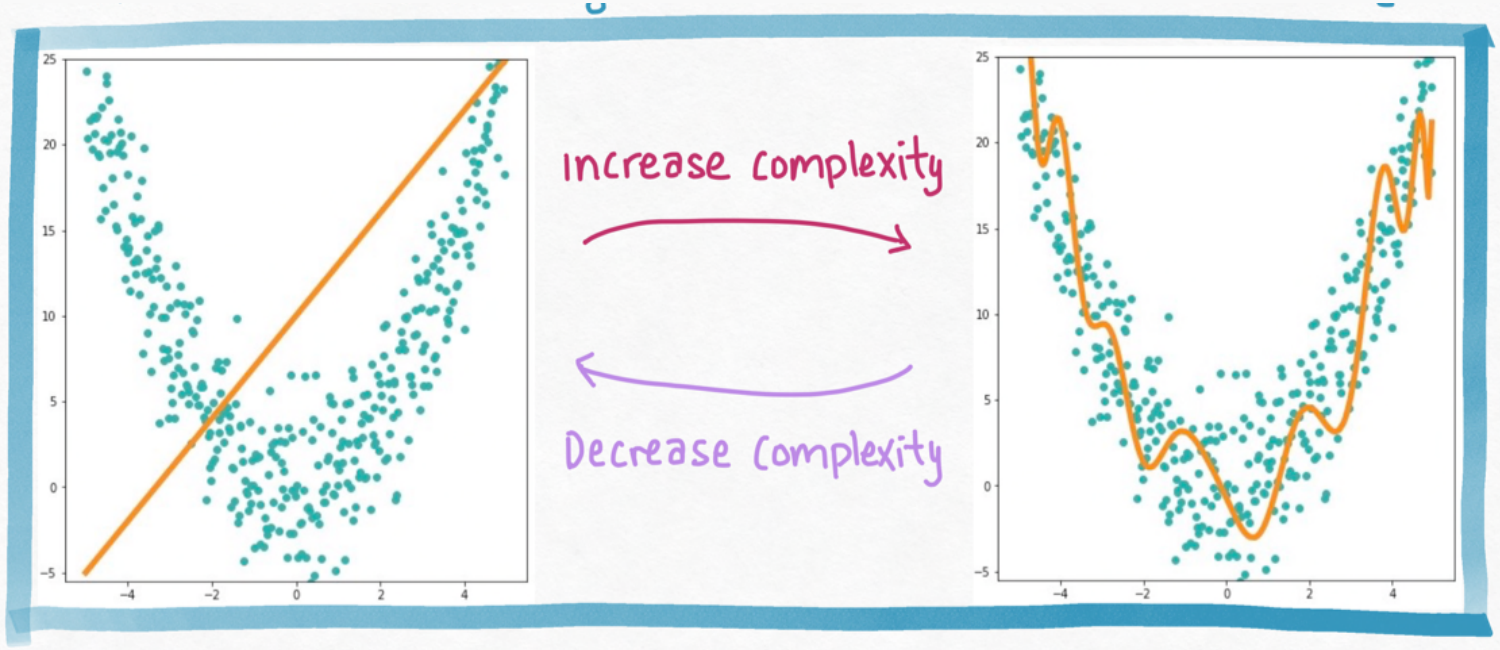
We distinguished the contributions of noise to the generalization error:

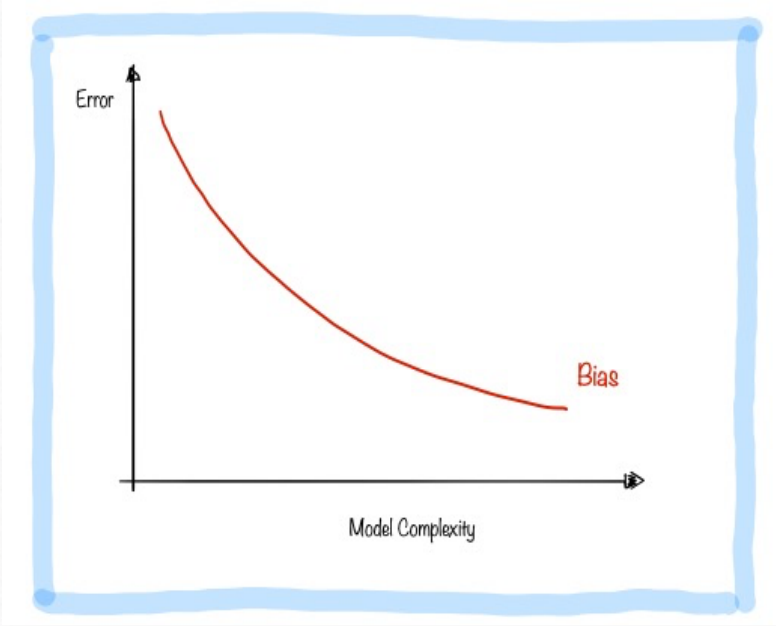
Irreducible error: we can't do anything to decrease error due to noise.

Reducible error: we can decrease error due to overfitting and underfitting by improving the model.

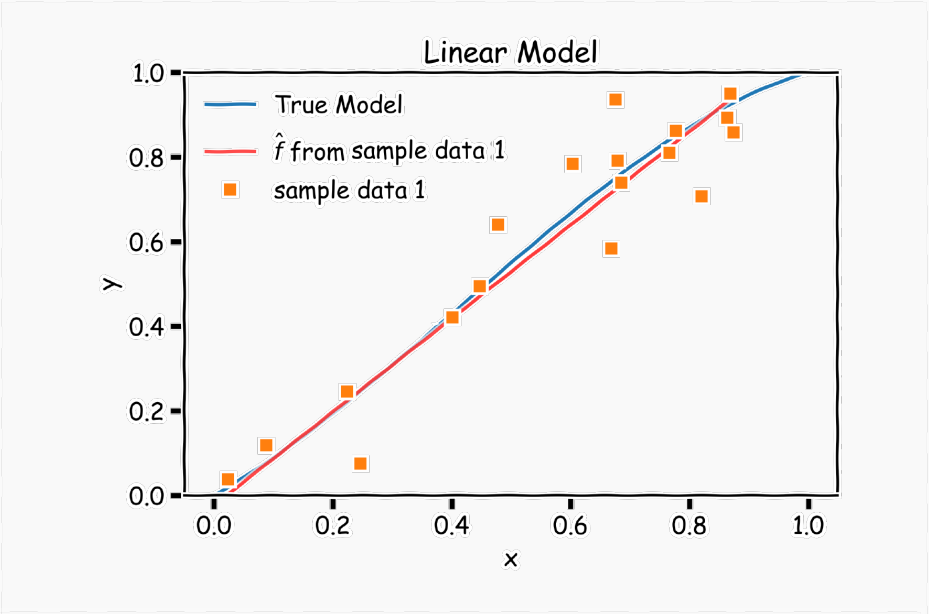
The Bias-Variance: Bias

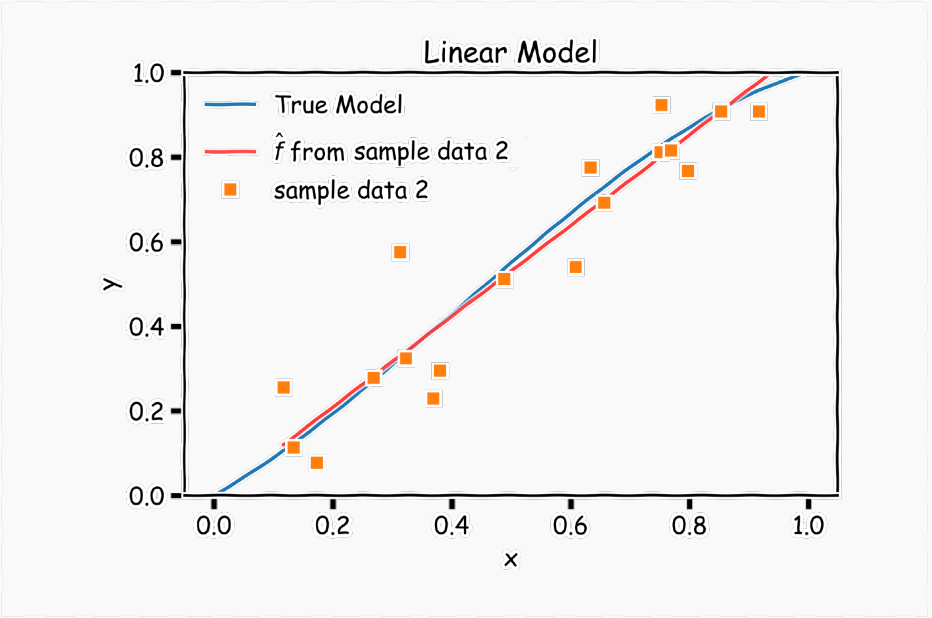
Reducible error comes from either underfitting or overfitting. There is a trade-off between the two sources of errors:

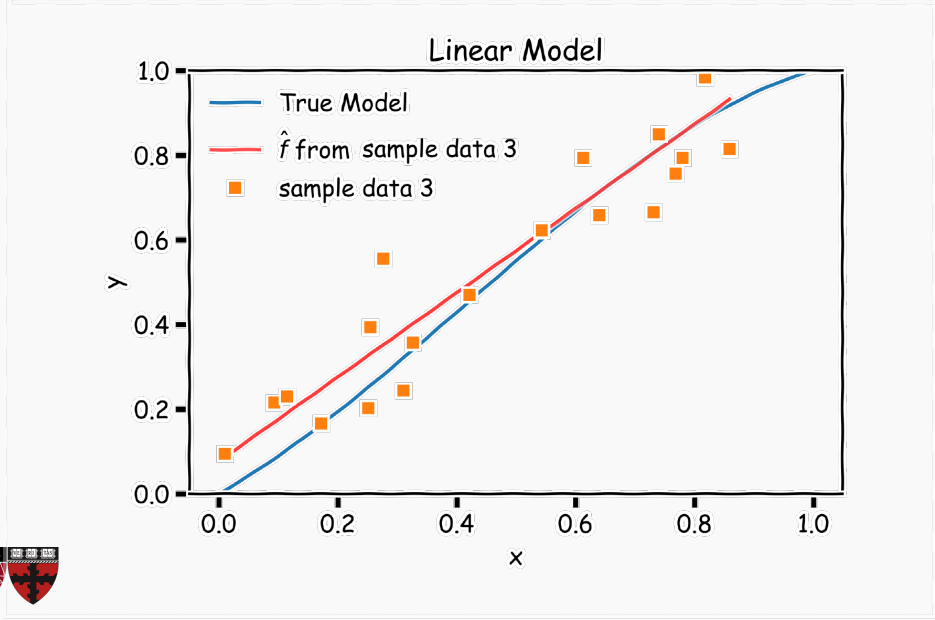


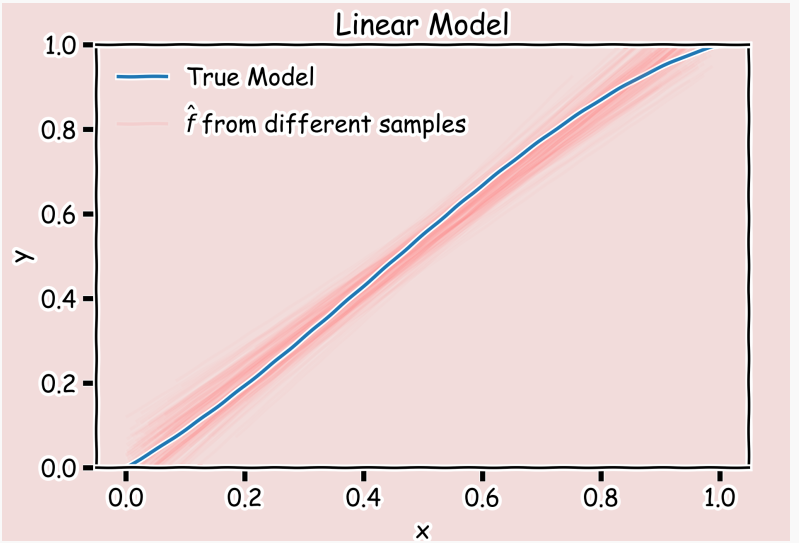


Bias vs Variance: Variance

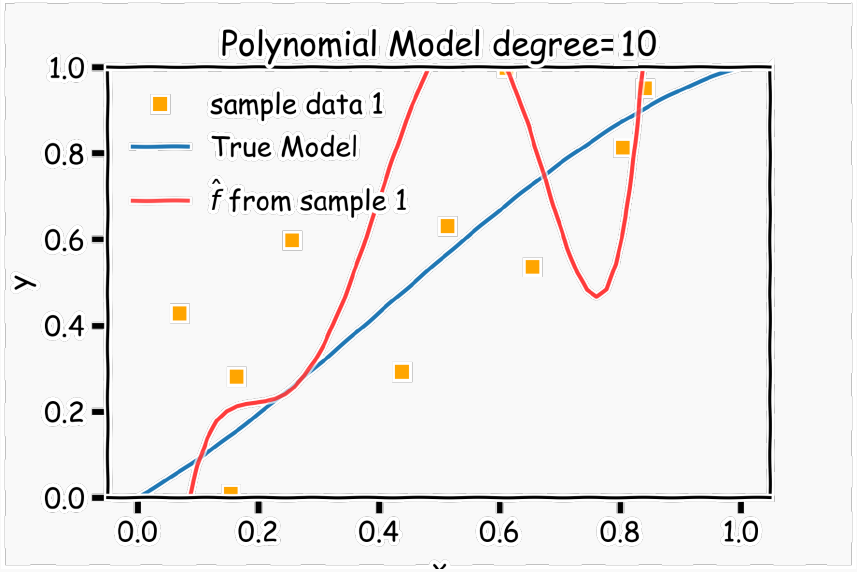


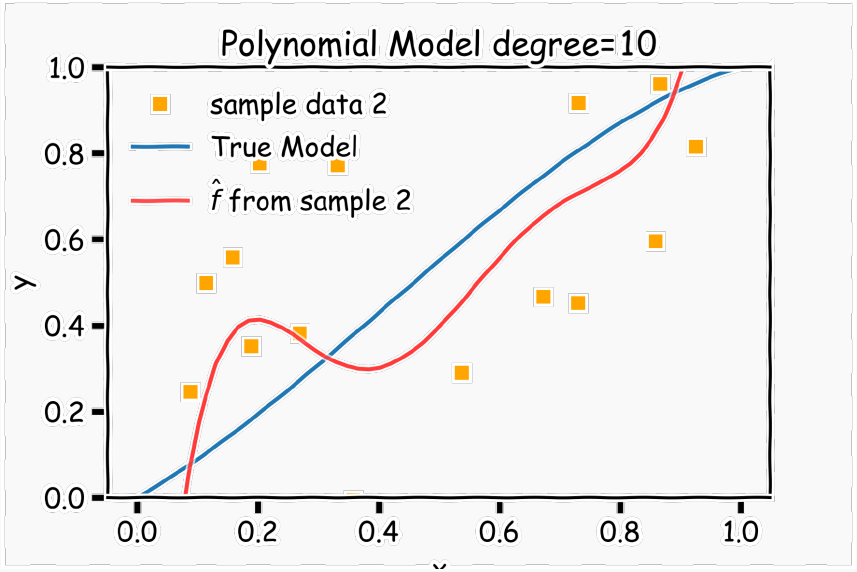


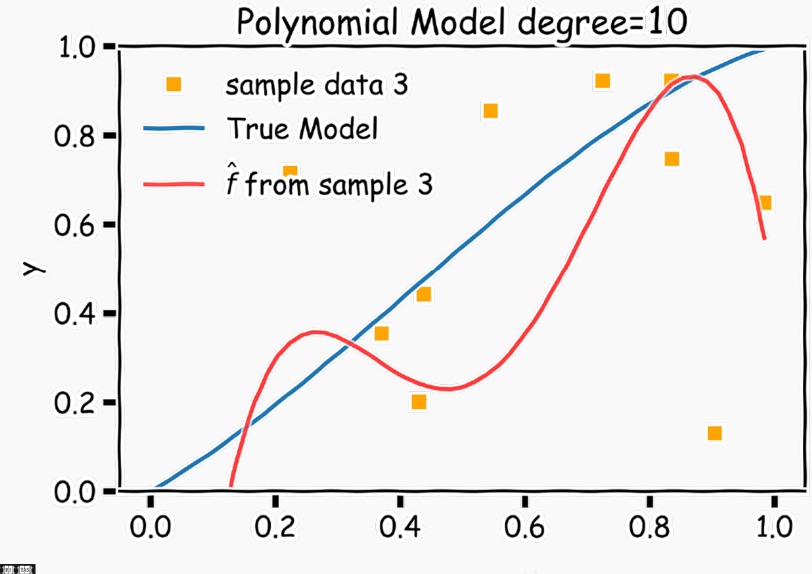


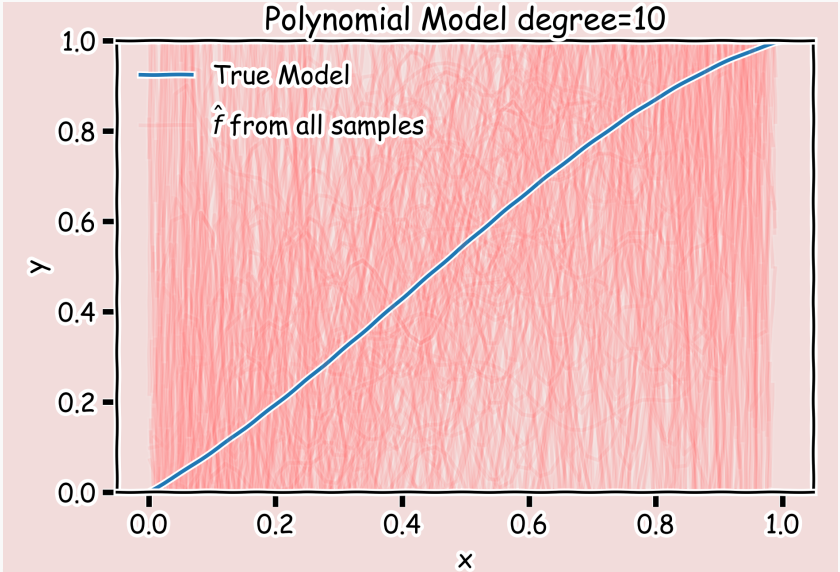


Bias vs Variance





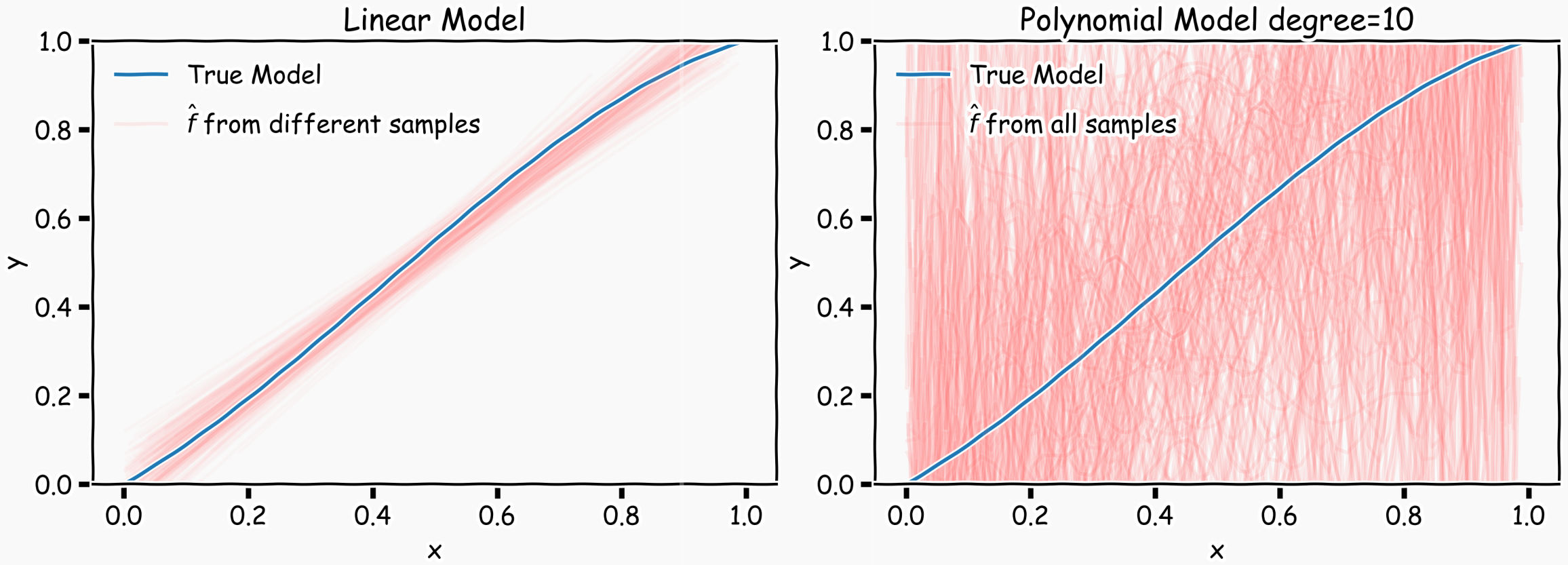




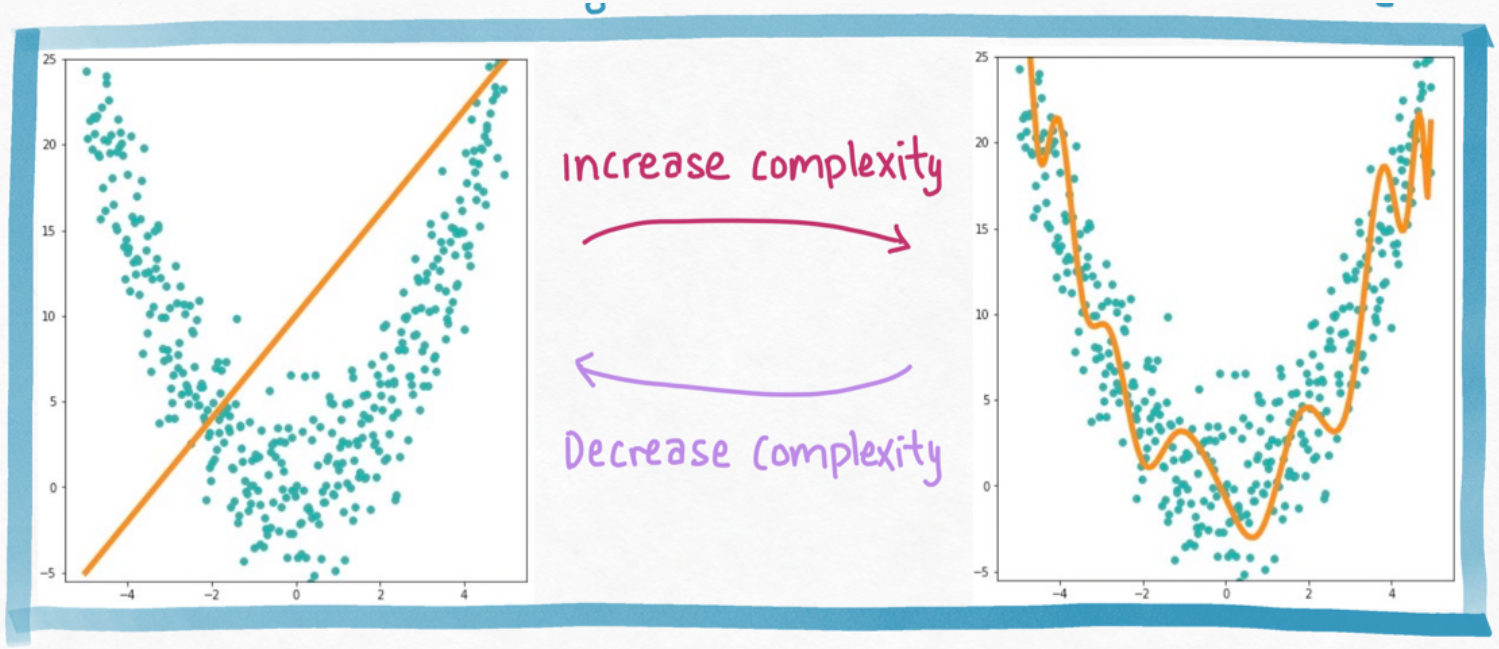
Bias vs Variance

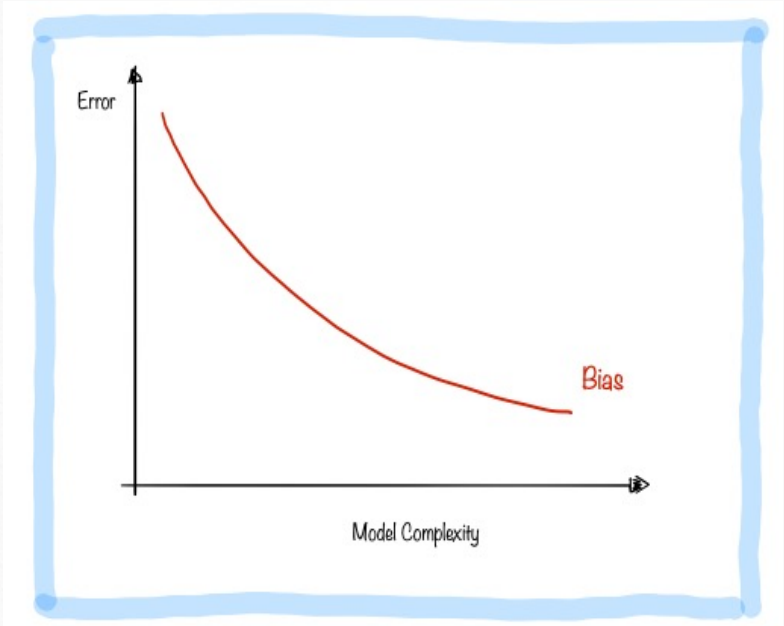
Left: 2000 best fit straight lines, each fitted on a different 20-points training set.

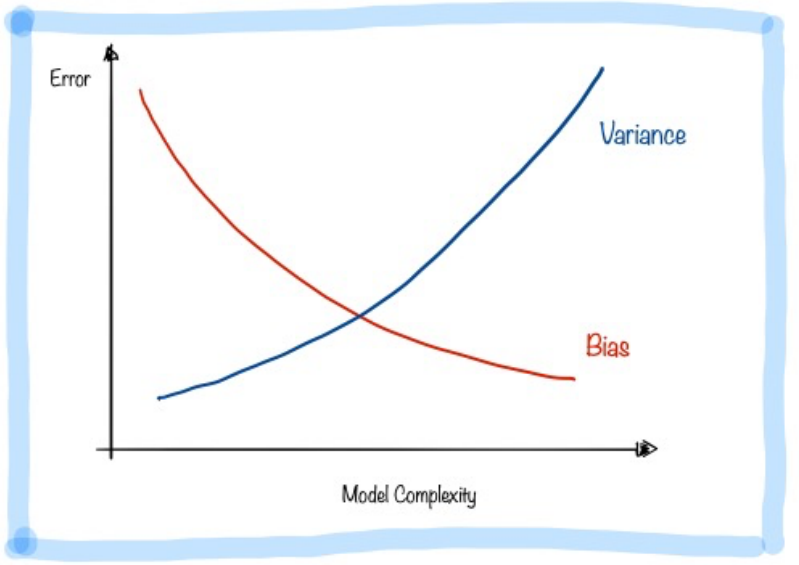
Right: Best-fit models using degree 10 polynomials.

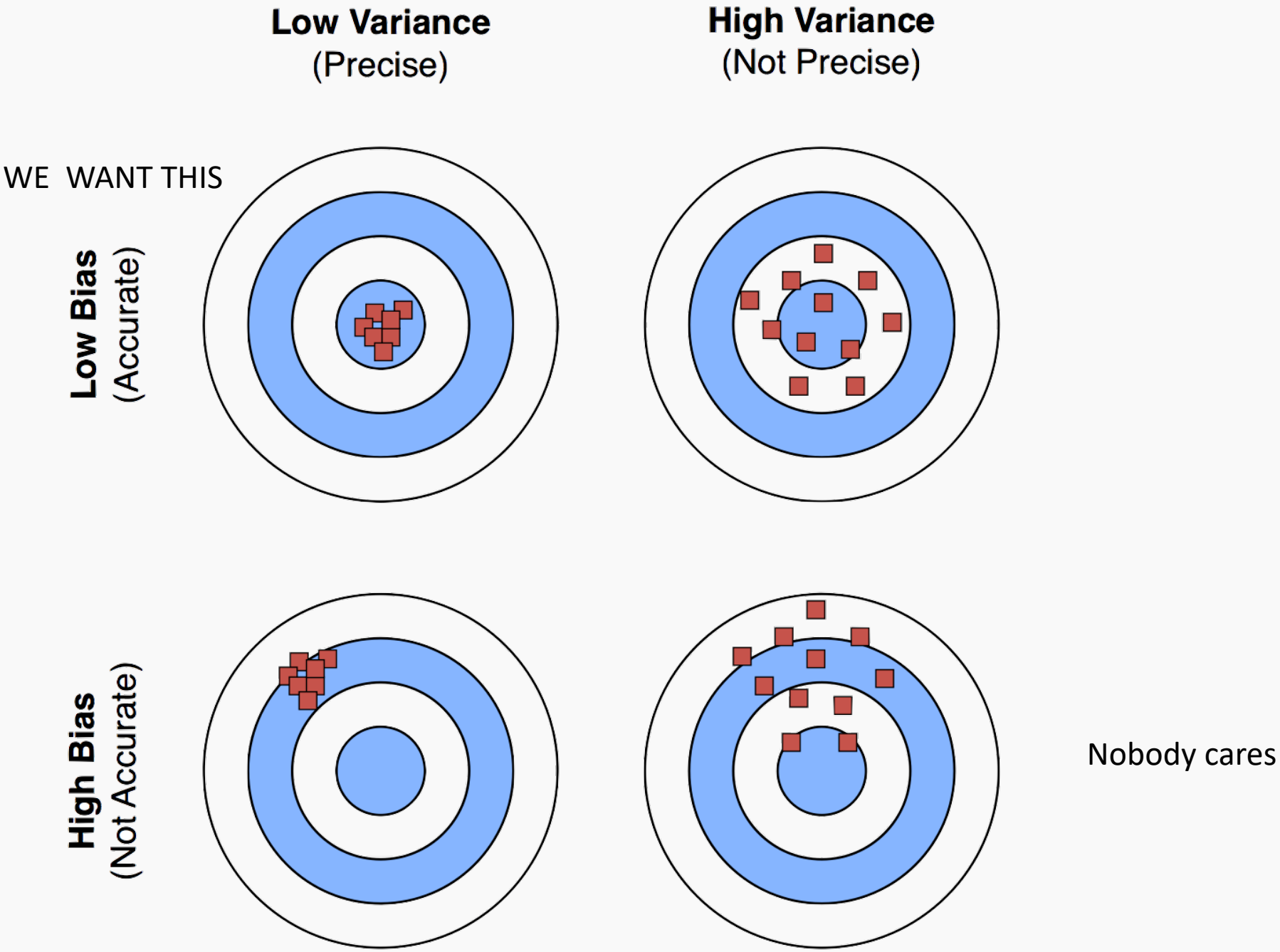


The Bias-Variance Trade Off: Bias









Nobody cares

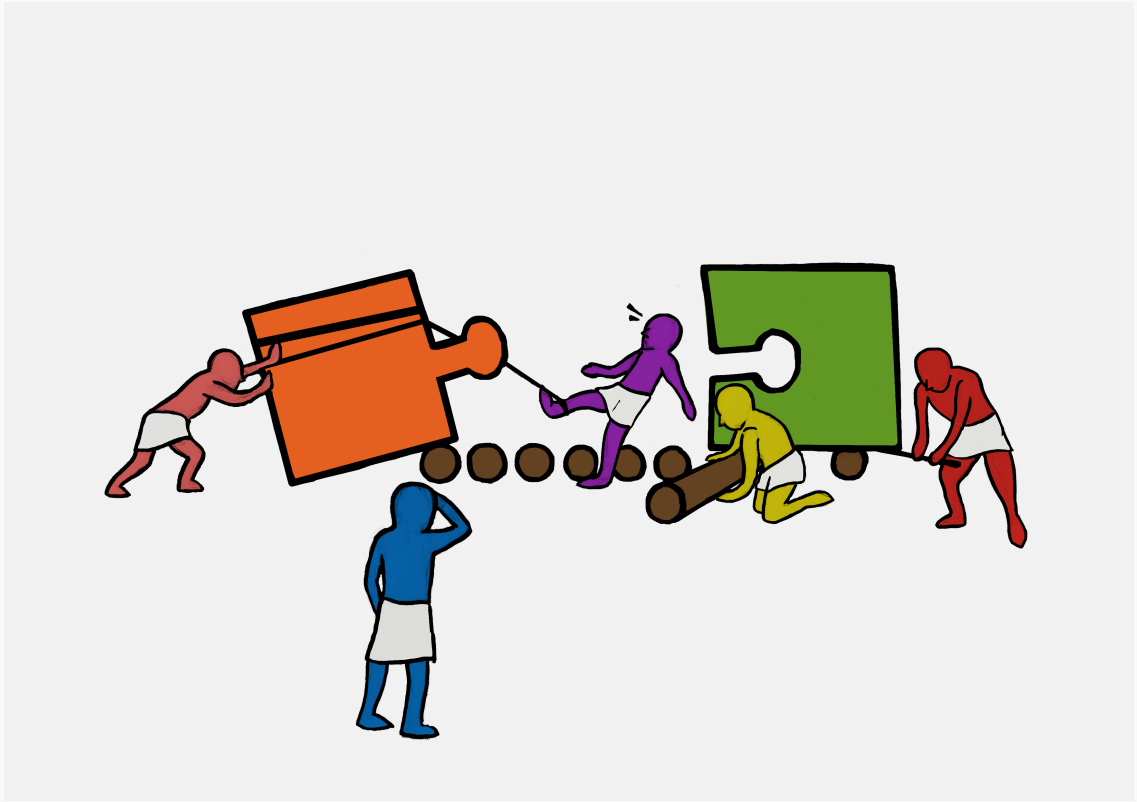
Overfitting

Overfitting occurs when a model corresponds too closely to the training set, and as a result, the model fails to fit additional data.

So far, we have seen that overfitting can happen when:

- Too many parameters
- Degree of the polynomial is too large
- Too many interaction terms

Next, we will see other evidence of overfitting, which will point to a way of avoiding overfitting: Ridge and Lasso regressions.



What to do? 🤔

- 👉 Today's lucky student: The first student in the breakout room
- 👉 **Instructions (a reminder):**
  - 👉 Listen to your peers' opinions and suggestions. Ask questions of each other ("What do you think"). Do not just lead others in the room without including everyone.
  - 👉 Make sure you do not cut-off or ignore what other students are trying to contribute.
- 👉 If you have questions, please reach out to the teaching staff. You can buzz us to come help, or if all else fails, come to the main room.