MMA/MMAI 869
Machine Learning and AI

Bias and Variance

Stephen Thomas

Updated: Oct 31, 2022

Smith SCHOOL OF BUSINESS | Queen's University

Outline

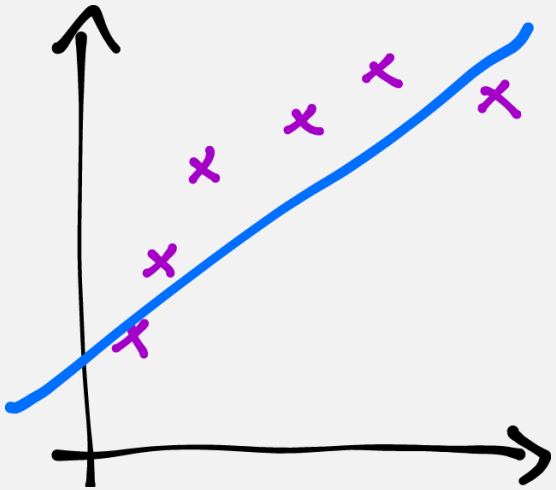
- What is bias and variance and why does it matter?

BIAS AND VARIANCE

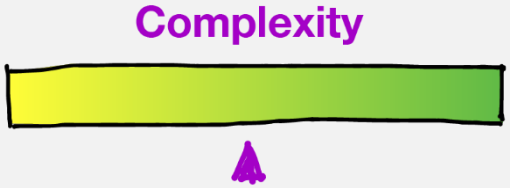
Bias/Variance Tradeoff

- Want model to not miss anything; capture all the interesting trends
- Don't want the model to be crazy, and risk **over-interpreting every outlier and irregularity**
- **It's a tradeoff**

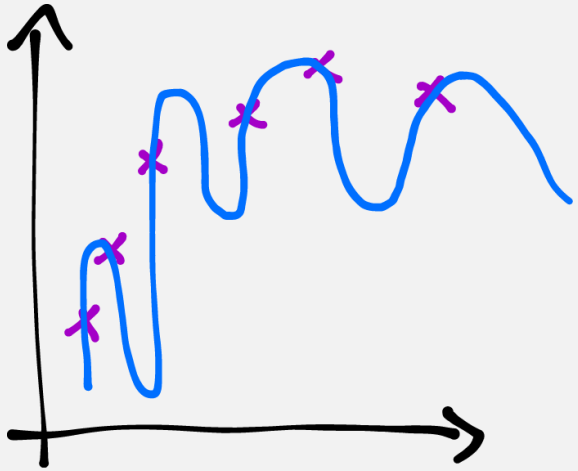
High Bias = Underfitting



- Model is too simple
- High bias -> Unable to capture true relationship; performance will be not great



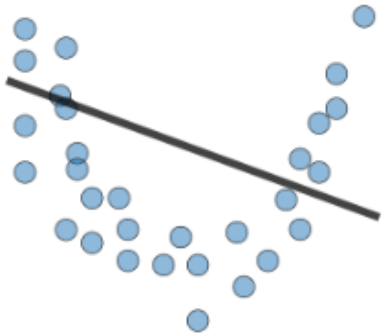
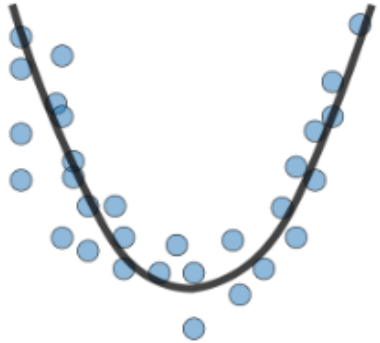
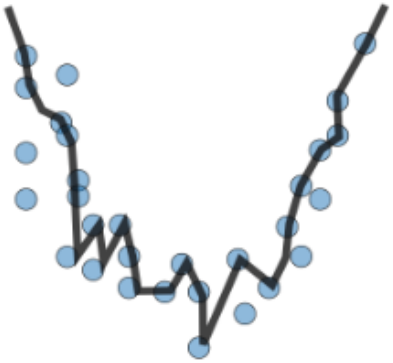
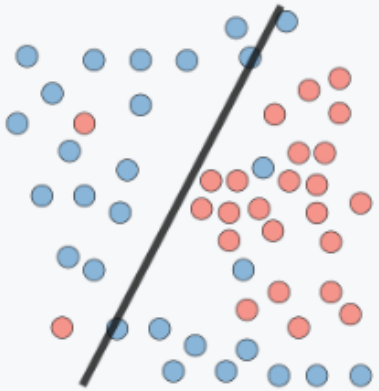
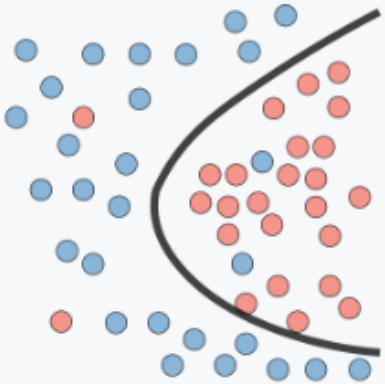
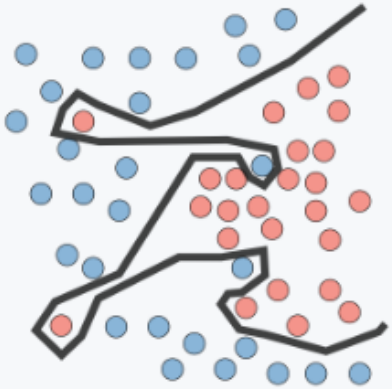
High Variance = Overfitting



- Model is too complex
- High variance -> Captures noise and outliers; performance will not be great

https://ml.berkeley.edu/blog/2017/07/13/tutorial-4/

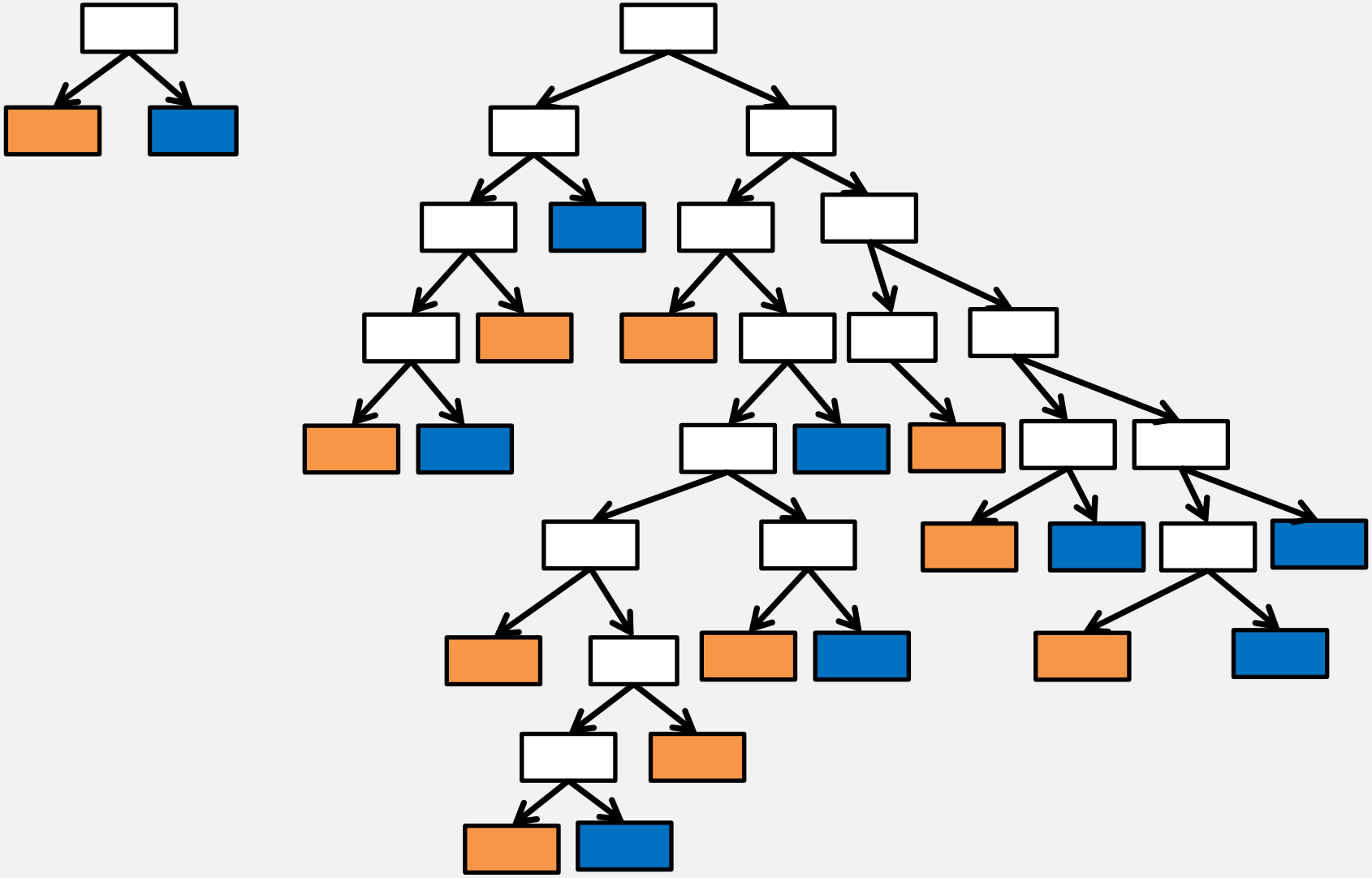
Regression Example

	Underfitting	Just right	Overfitting
Symptoms	• High training error • Training error close to test error • High bias	• Training error slightly lower than test error	• Very low training error • Training error much lower than test error • High variance
Regression illustration			
Classification illustration			
Possible remedies	• Complexify model • Add more features • Train longer		• Perform regularization • Get more data

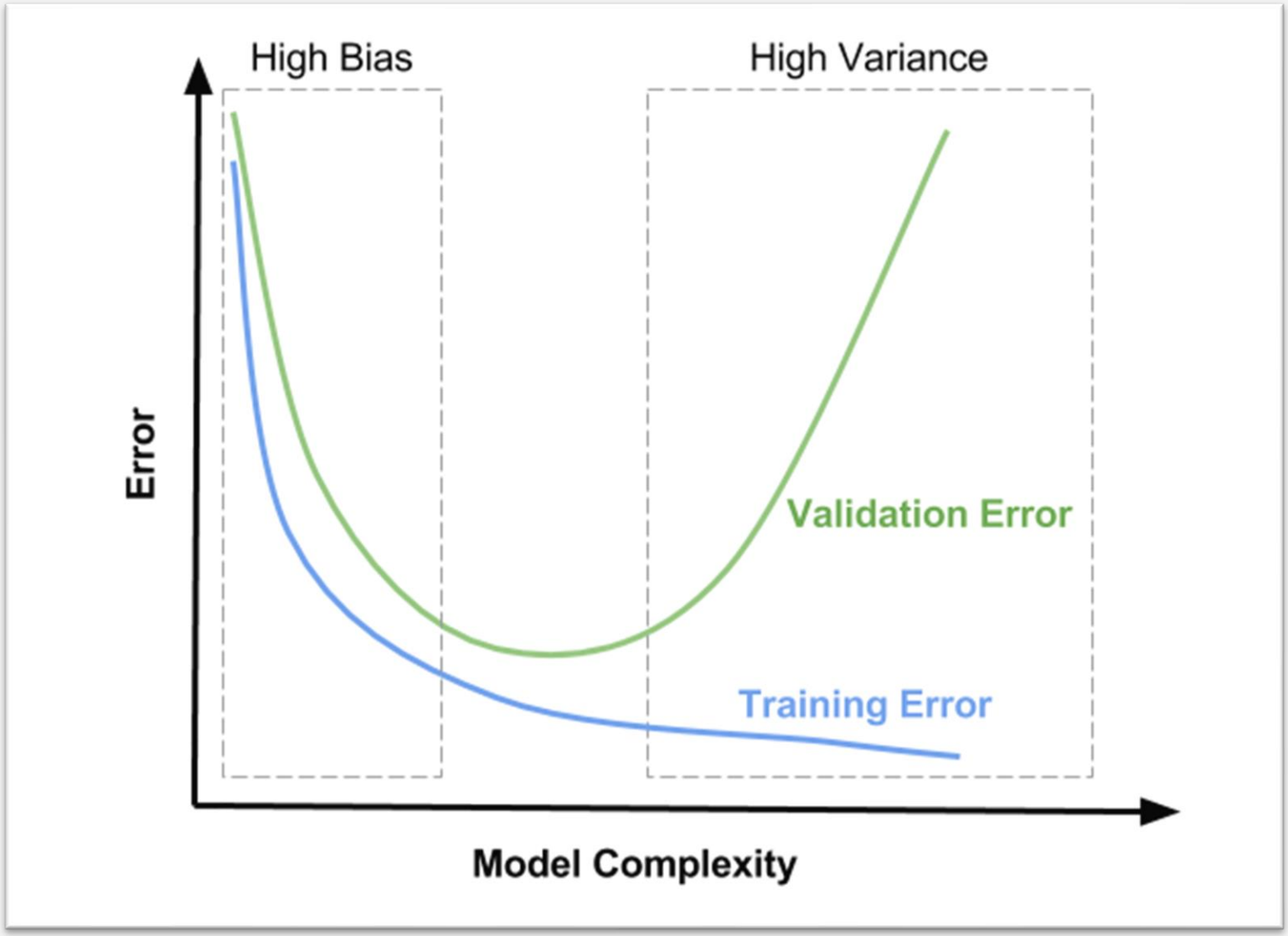
Stanford CS 229

Decision Tree Example

- Which tree has high bias, and which has high variance?



How to Tell?



https://www.learnopencv.com/bias-variance-tradeoff-in-machine-learning/

SUMMARY

Summary

- **High Bias** = Underfitting the training data
  - Model won't be good enough
- **High Variance** = Overfitting the training data
  - Model won't be good enough
- Want somewhere in the middle