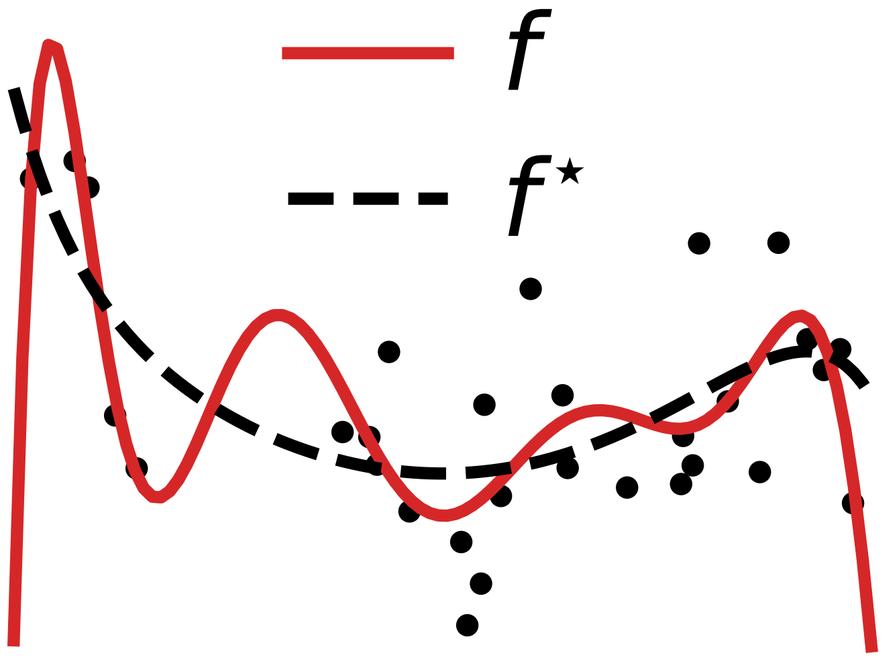
Bias and variance

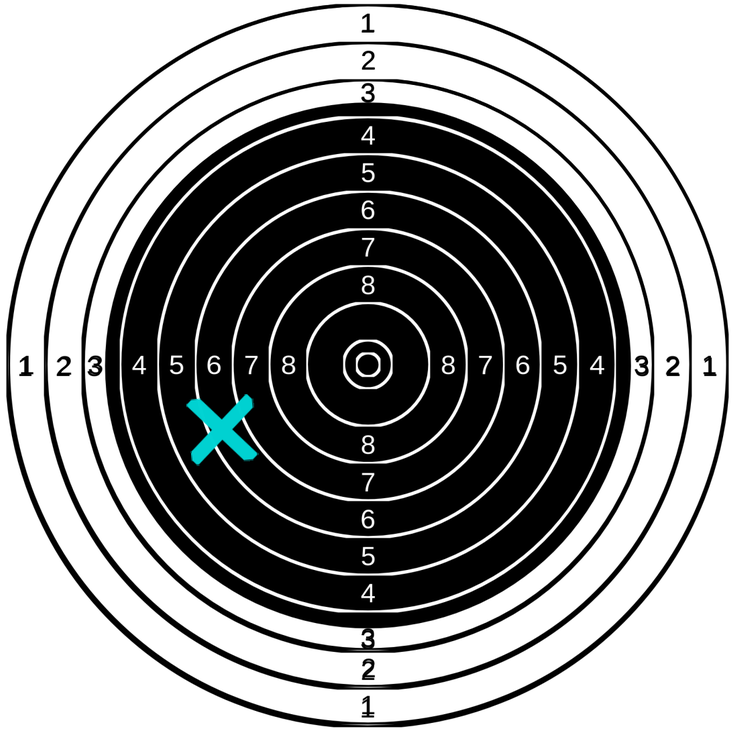
A statistical view of underfitting and overfitting

Resampling the training set

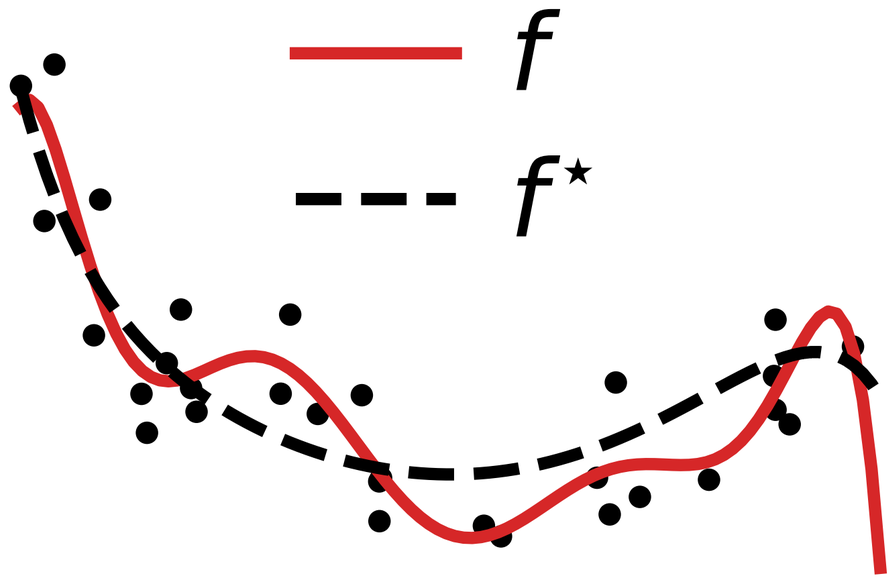
- A limited amount of training data
- Training set is a small random subset of all possible observations
- What is the impact of this choice of training set on the learned prediction function?

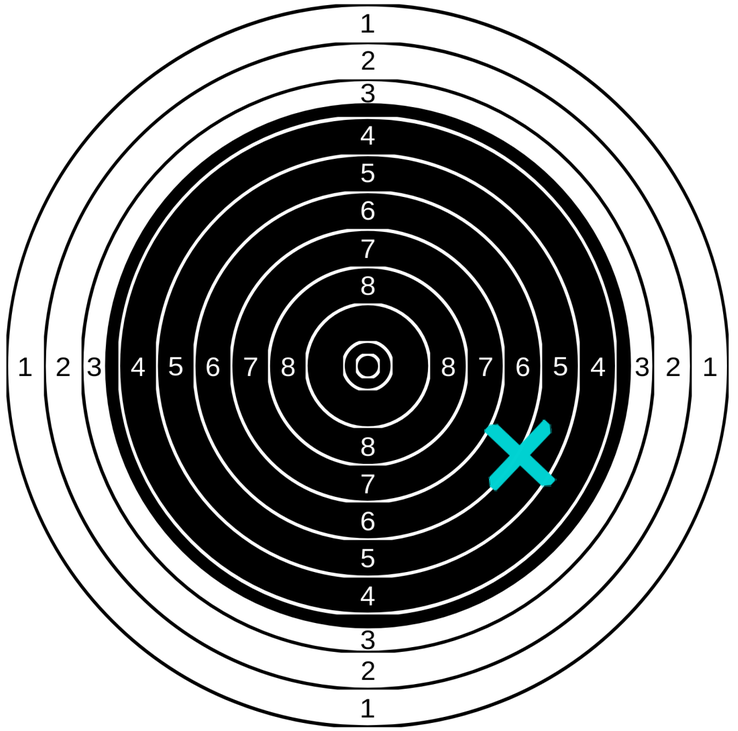
Overfit: variance



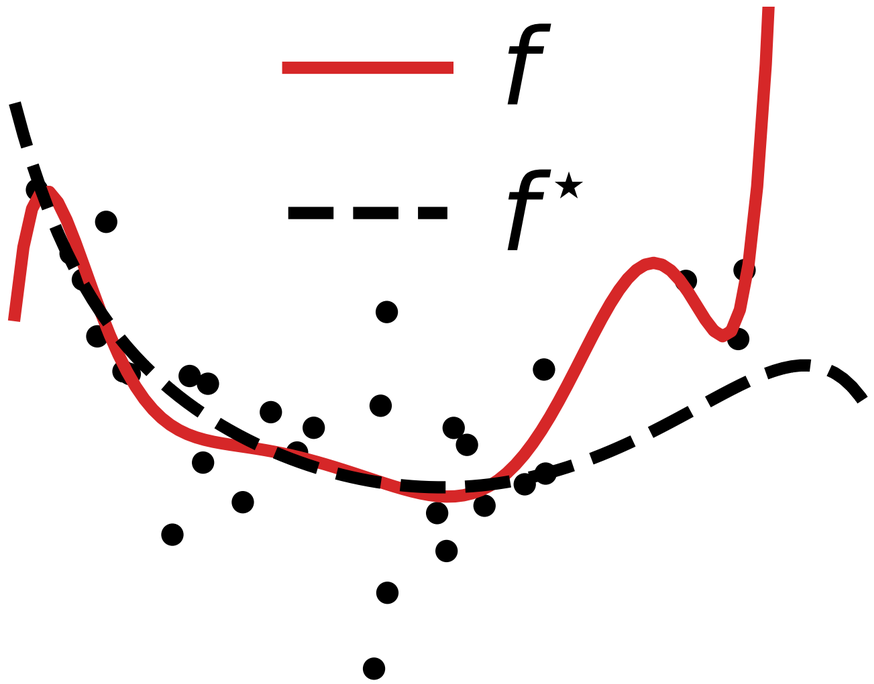


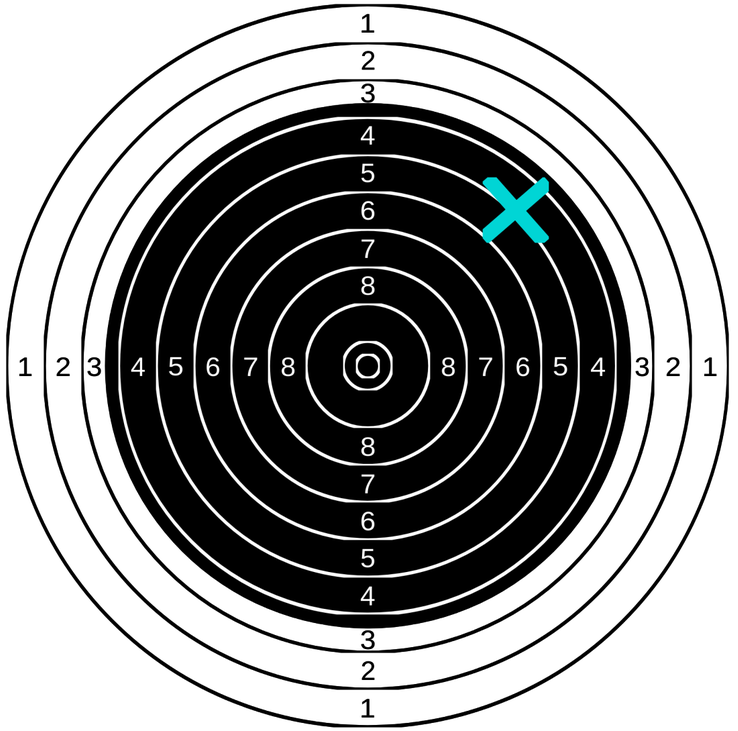
Overfit: variance



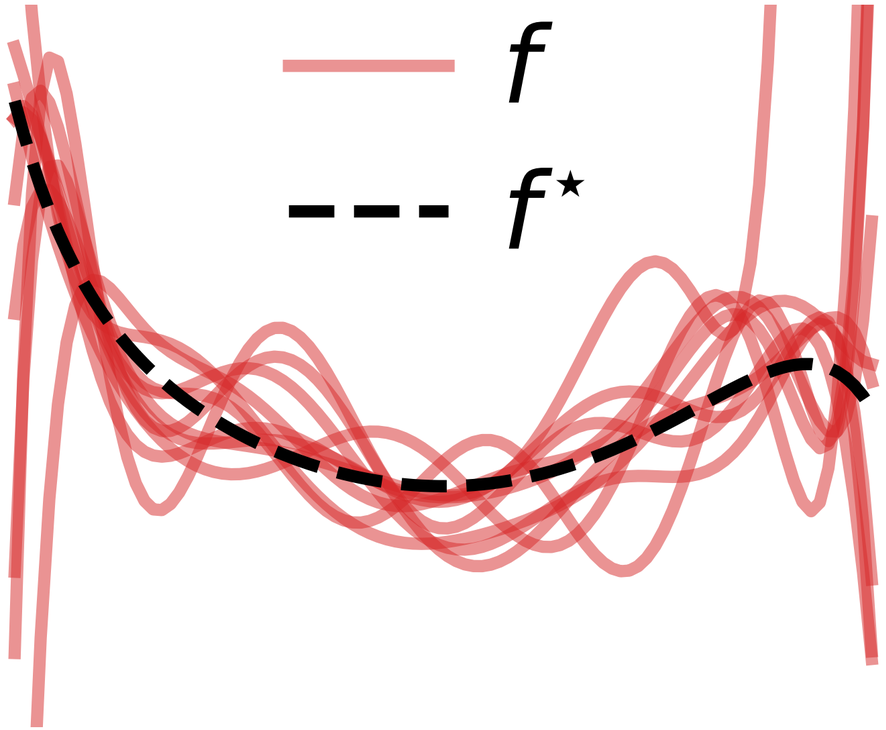


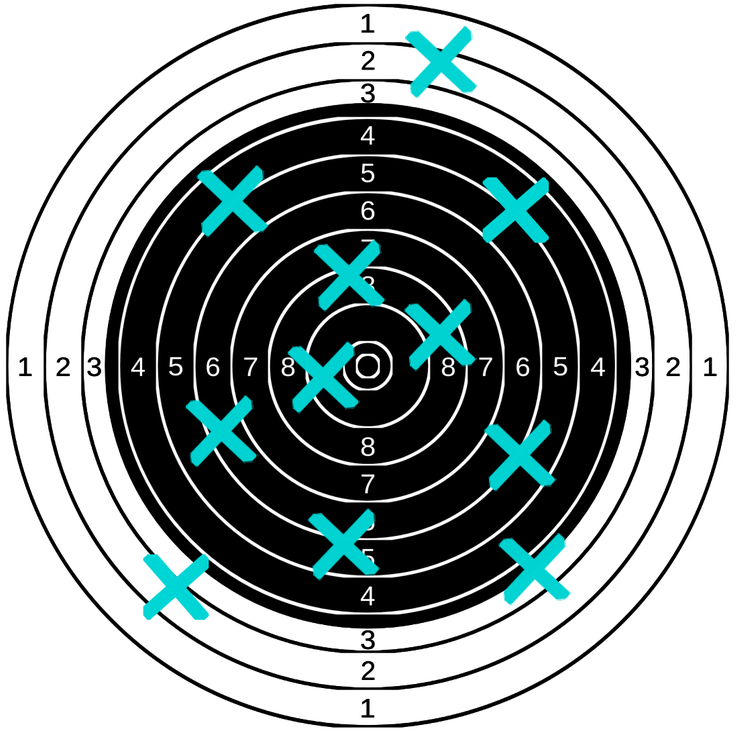
Overfit: variance



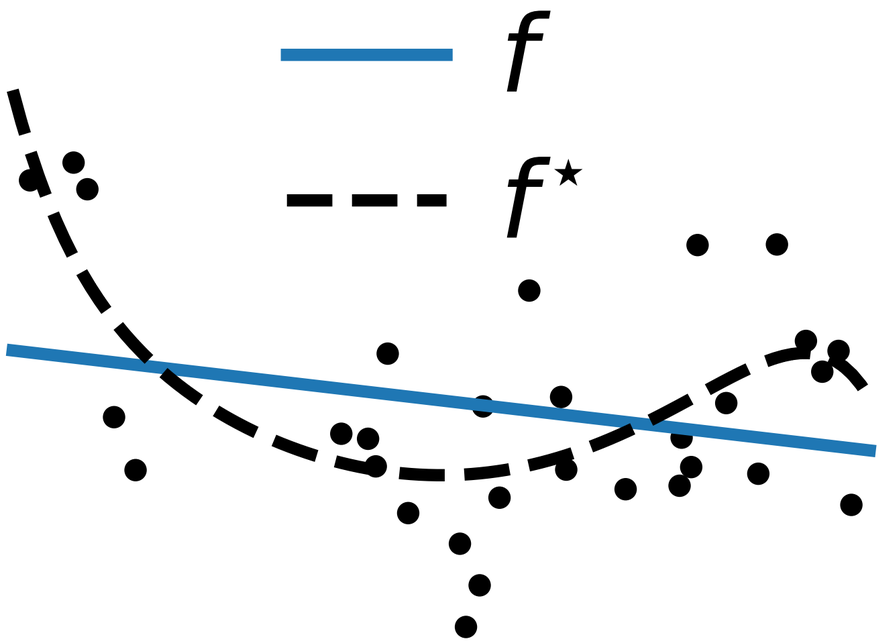


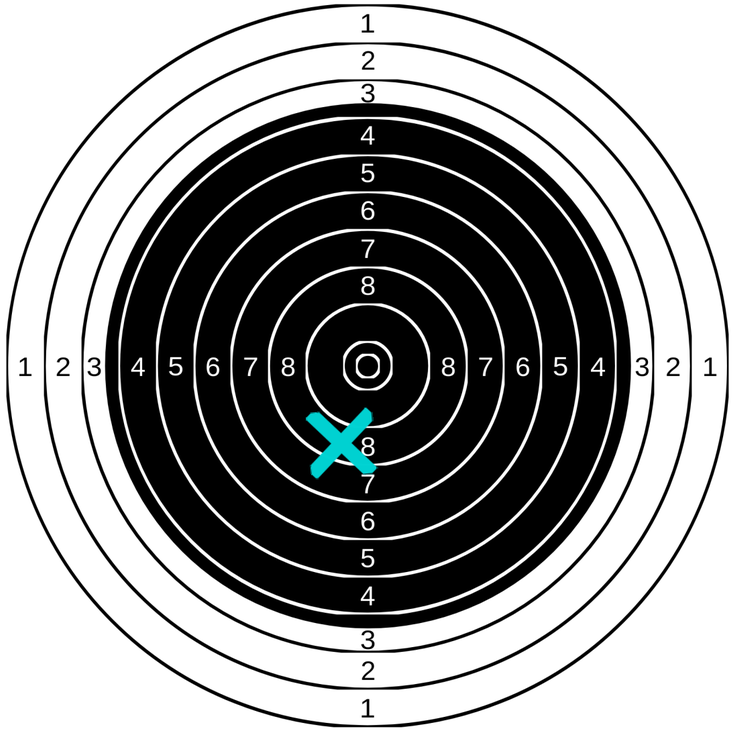
Overfit: variance



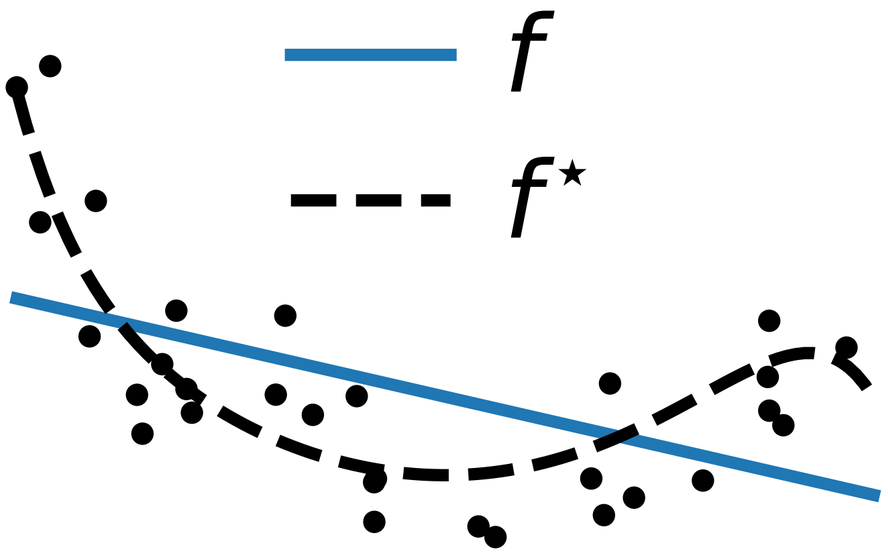


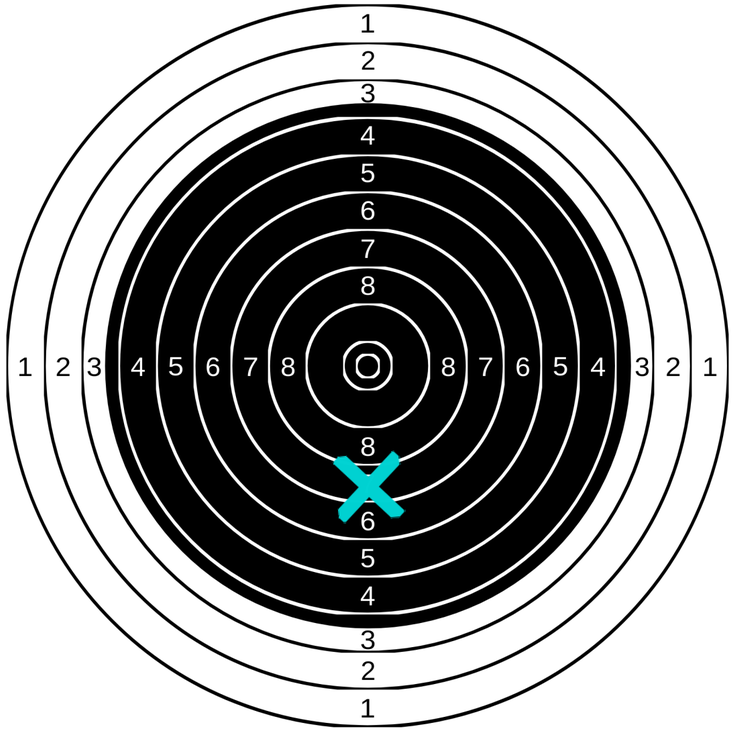
Underfit: bias



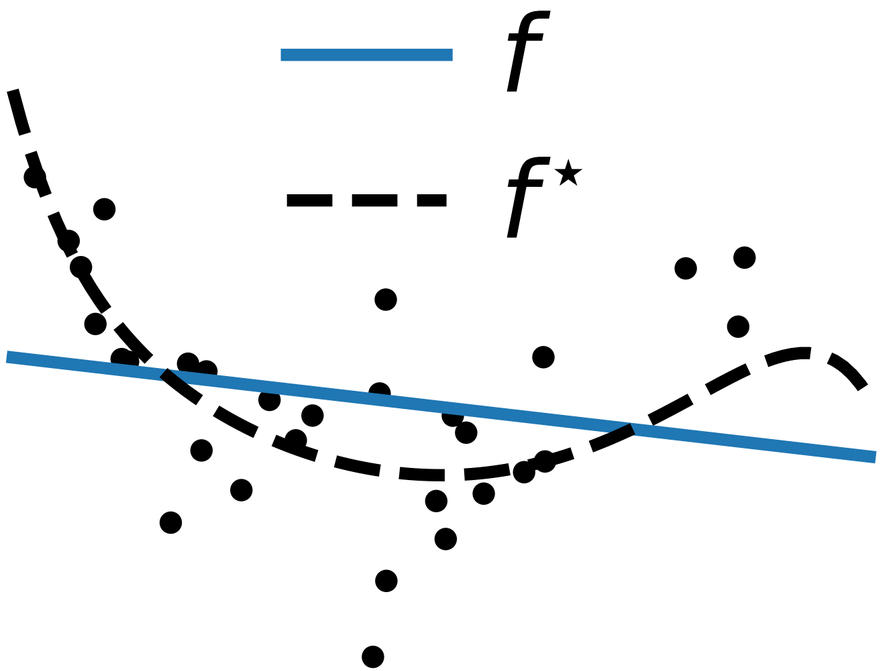


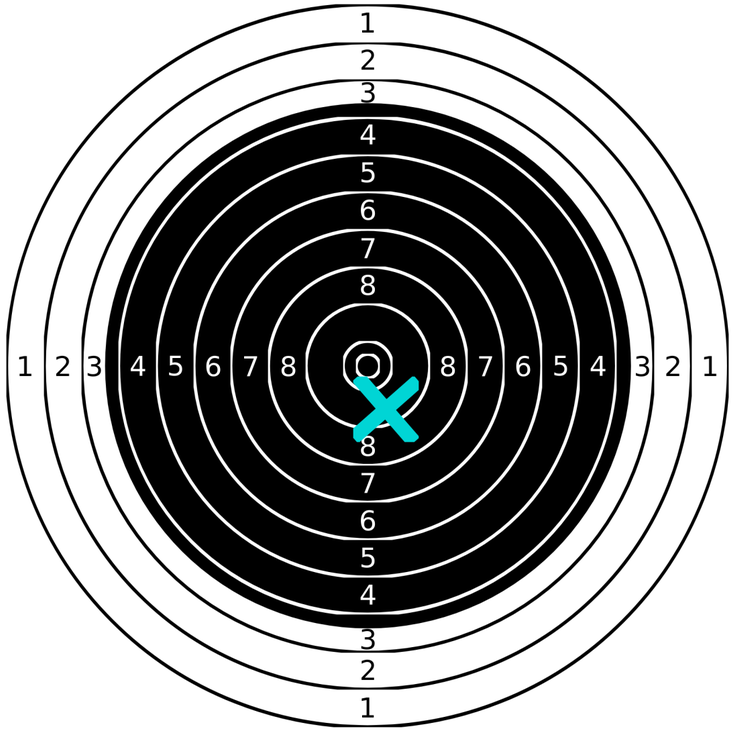
Underfit: bias



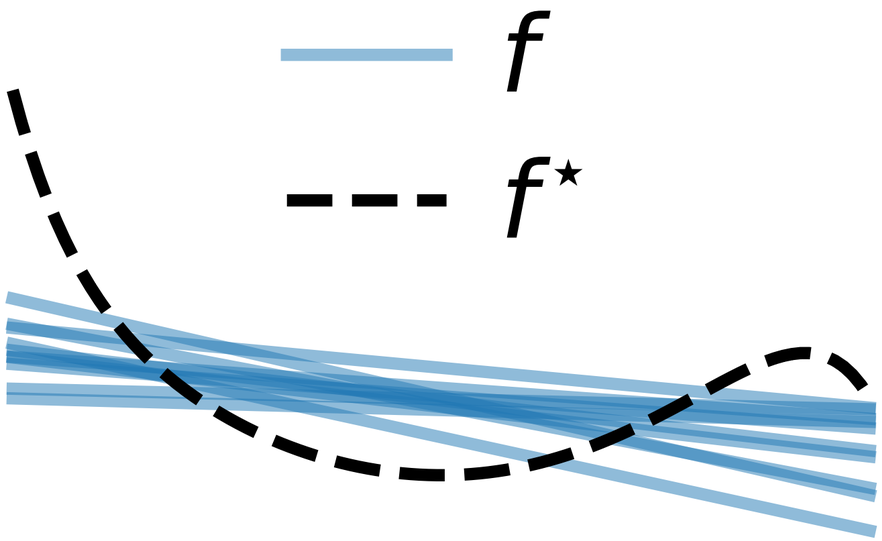


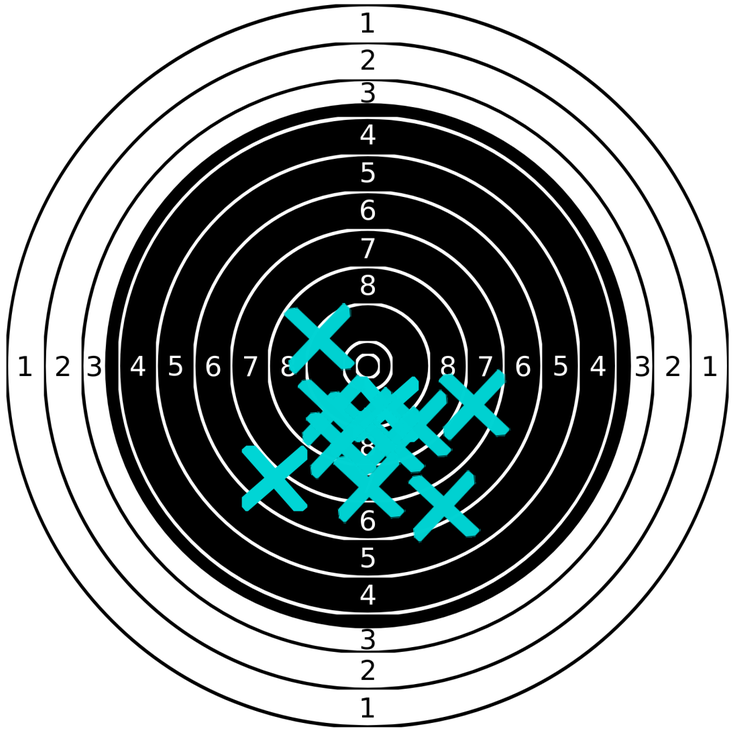
Underfit: bias



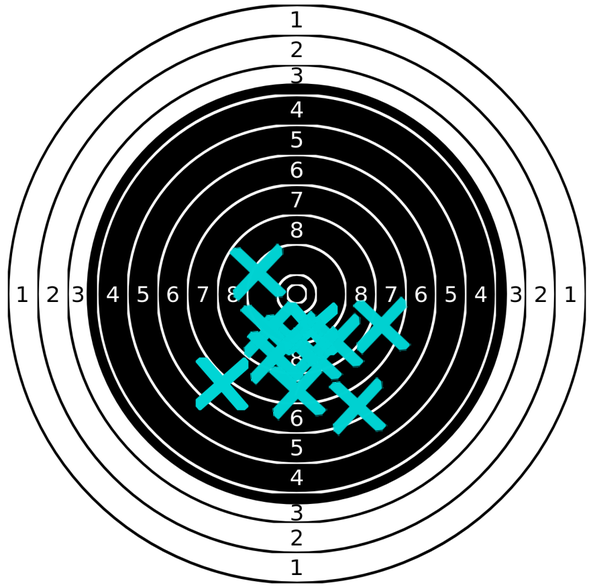


Underfit: bias

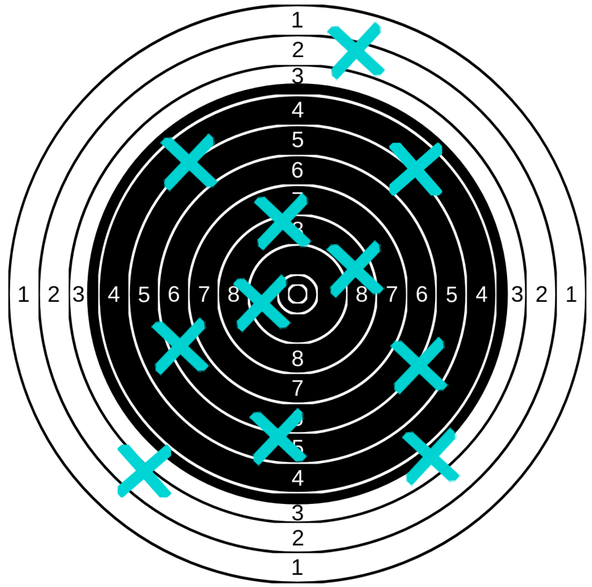




Underfit vs overfit



Bias



Variance

The bias-variance decomposition of the Mean Squared Error (MSE)

For people with a background in mathematics and statistics who are interested in a more formal treatment of those concepts:
Decomposition of the squared prediction error on Wikipedia

Note that the MOOC evaluation does not require you to understand the mathematical details, only the general intuition.

Main takeaways

High bias == underfitting:

- systematic prediction errors
- the model prefers to ignore some aspects of the data
- mispecified models

High variance == overfitting:

- prediction errors without obvious structure
- small change in the training set, large change in the model
- unstable models

The bias can come from the choice of the model family.