The Bias-Variance Tradeoff

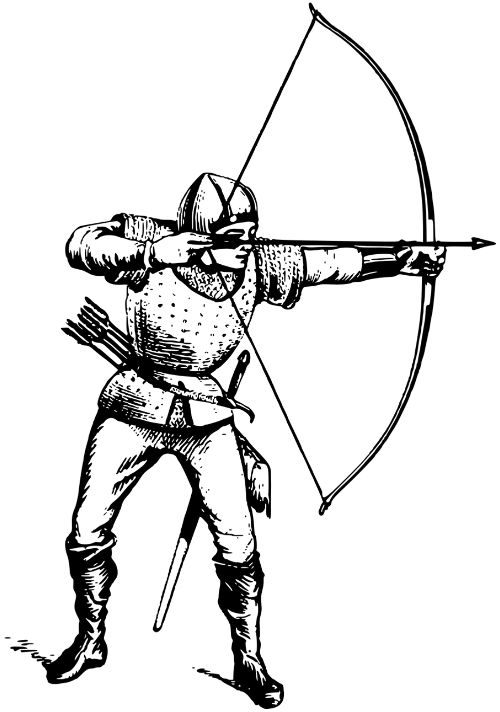
Learning Objectives

- Describe bias and variance in machine learning
- Describe what it means to underfit or overfit to data
- Relate underfitting and overfitting to model bias and variance
- Answer two very common interview questions:

What is the bias-variance trade-off?

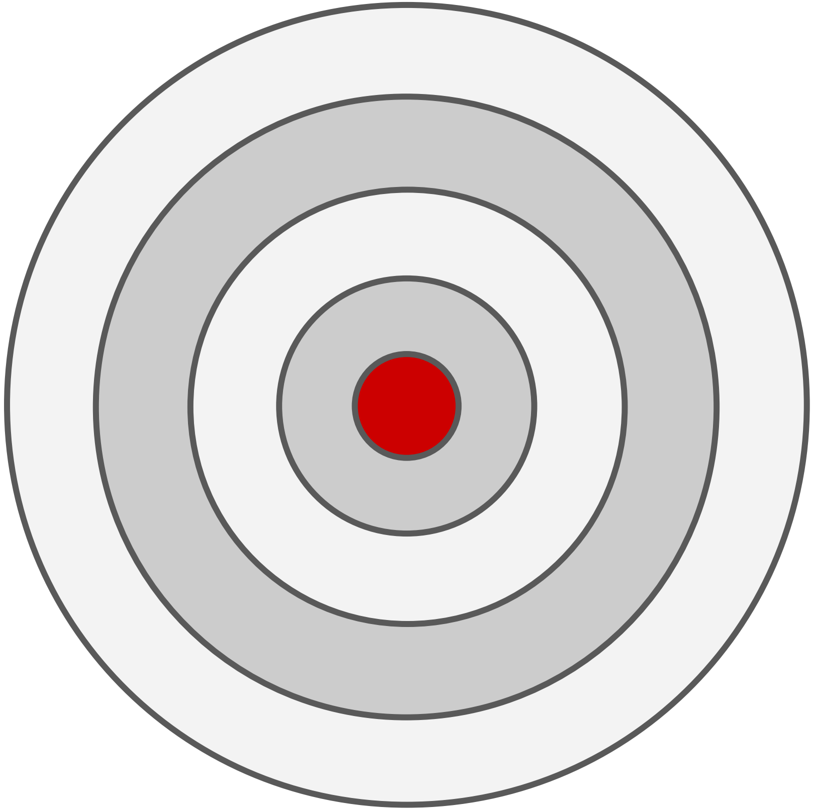
Why is there a bias-variance trade-off?

Bias and Variance Introduction



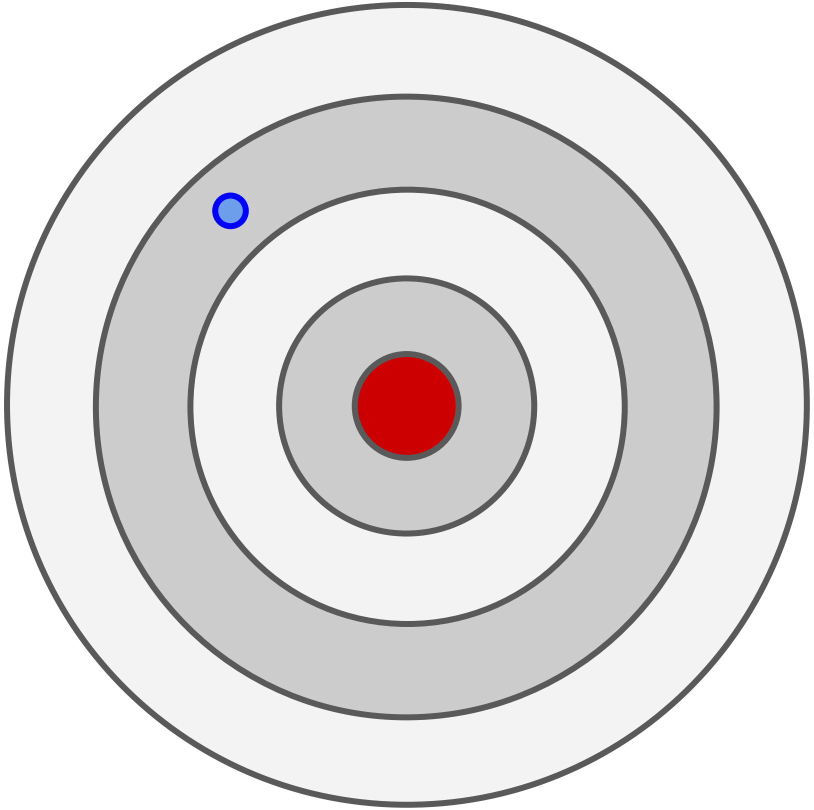
Bias and Variance Introduction

The archer takes aim and ...



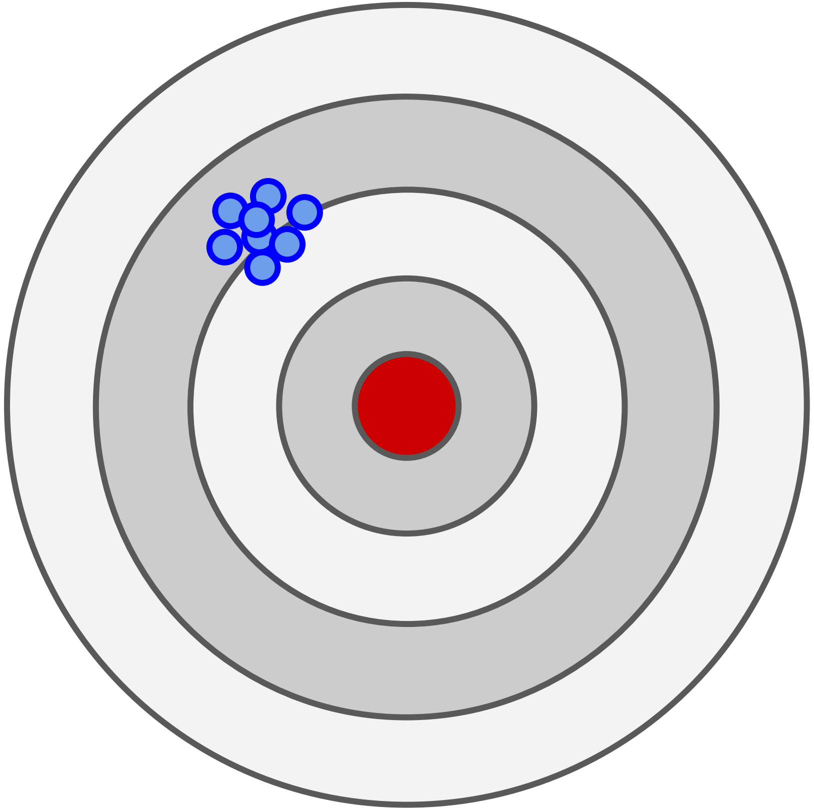
Bias and Variance Introduction

Releases...



Bias and Variance Introduction

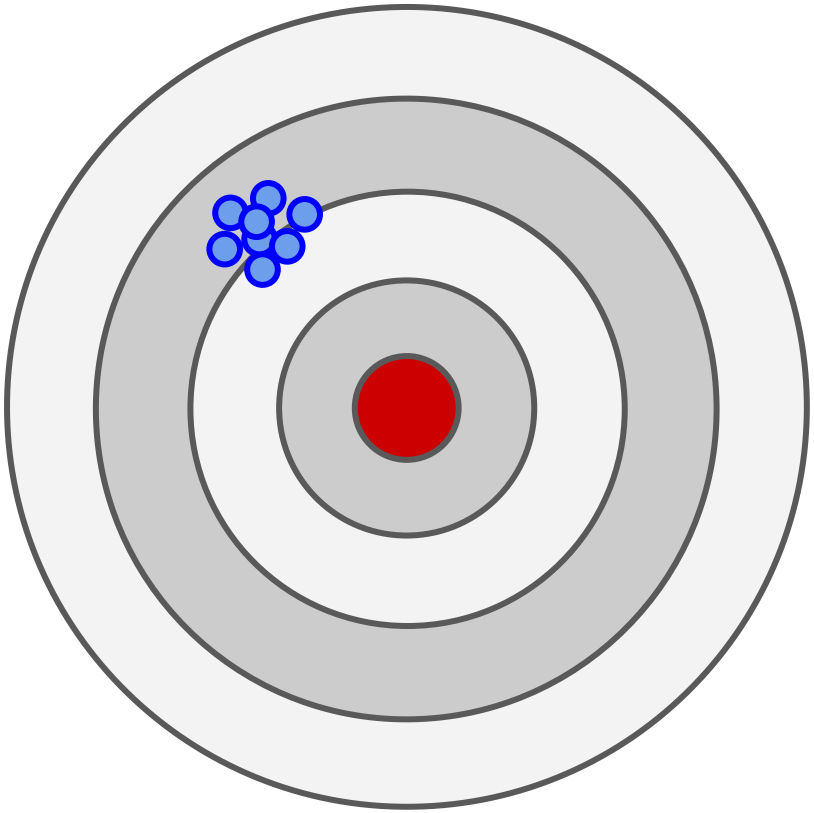
And then shoots 7 more ...



Bias and Variance Introduction

Statistical **bias**:

The amount the **expected value** of the results differ from the true value.



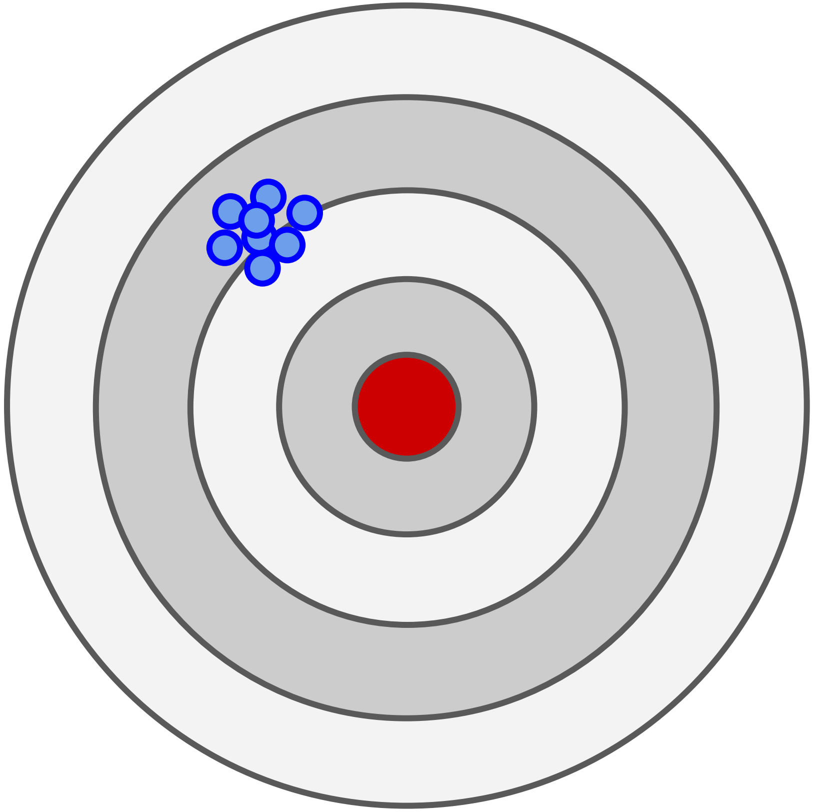
Bias and Variance Introduction

Statistical **bias**:

The amount the **expected value** of the results differ from the true value.

Expected value:

The **long-term average** value of the results of an experiment or function.



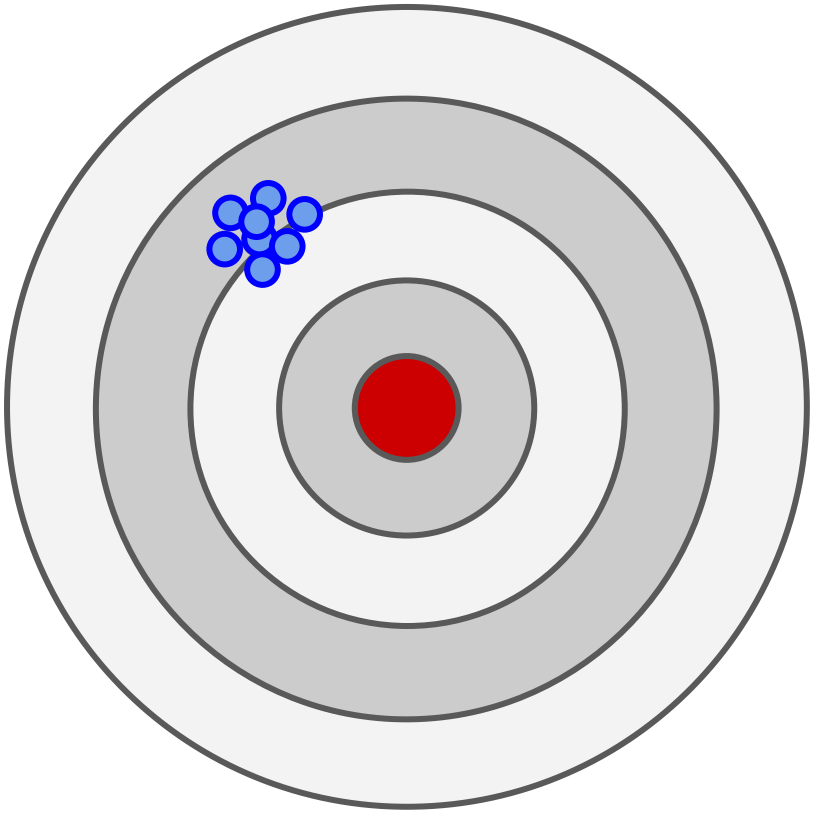
Bias and Variance Introduction

So in this archery analogy:

What is the “true value”?

What are the “results”?

What’s the model?



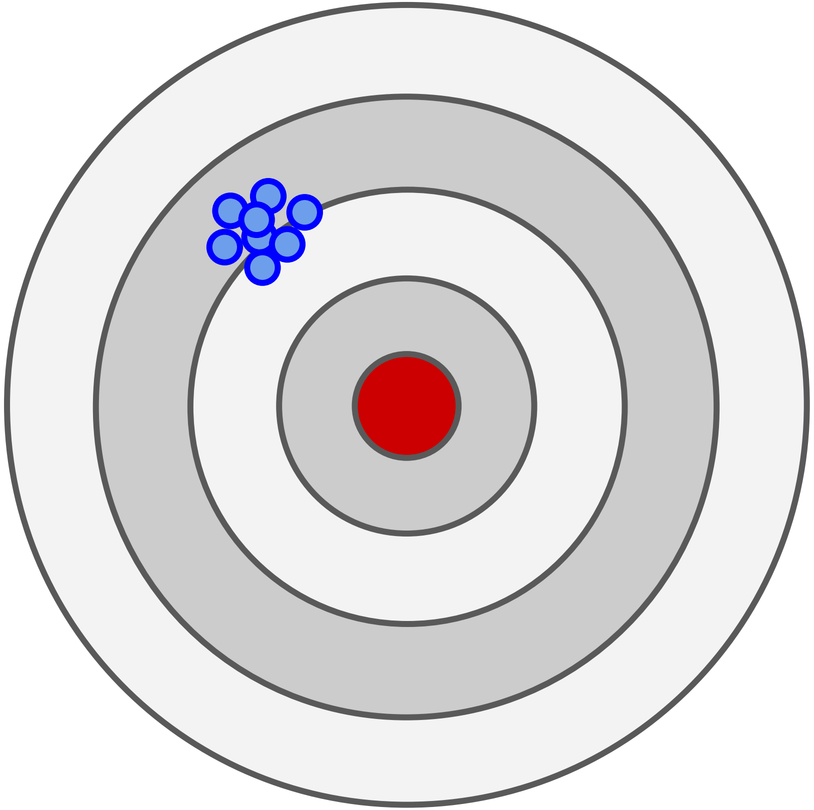
Bias and Variance Introduction

So in this archery analogy:

What is the “true value”?
The bullseye.

What are the “results”?
Where the arrows landed.

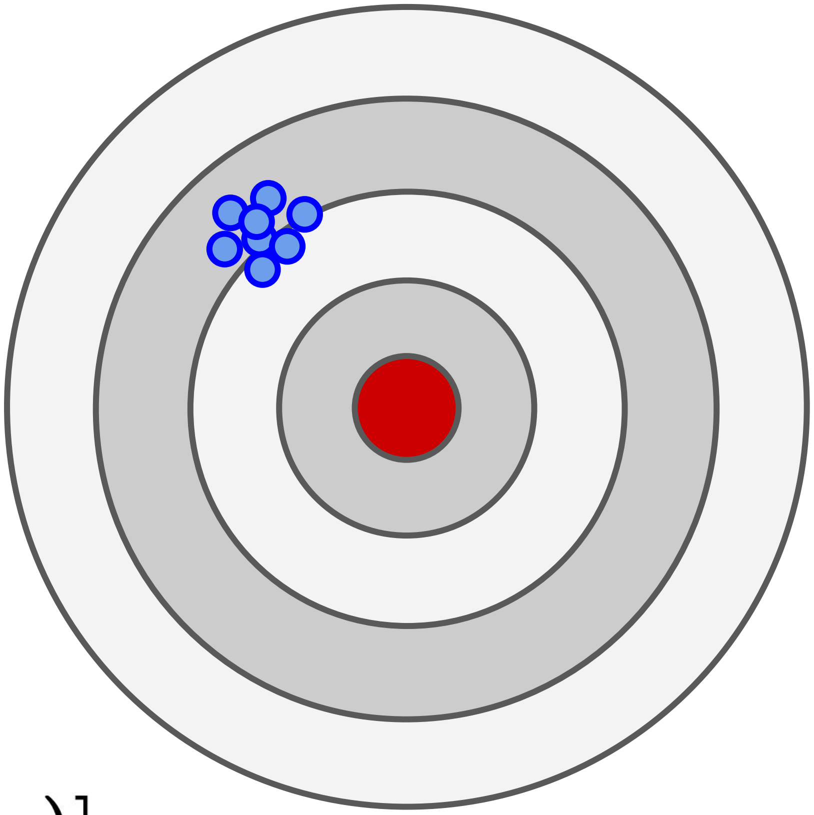
What’s the model?
The archer.



Bias and Variance Introduction

So in this archery analogy:

- What is the "true value"?
  the bullseye
- What are the "results"?
  where the arrows landed
- What's the model?
  the archer



Adding some statistical notation:

true value: $f(x)$

results: $\hat{f}(x)$

the expected value of the results: $\mathbf{E}[\hat{f}(x)]$

Bias and Variance Introduction

So in this archery analogy:

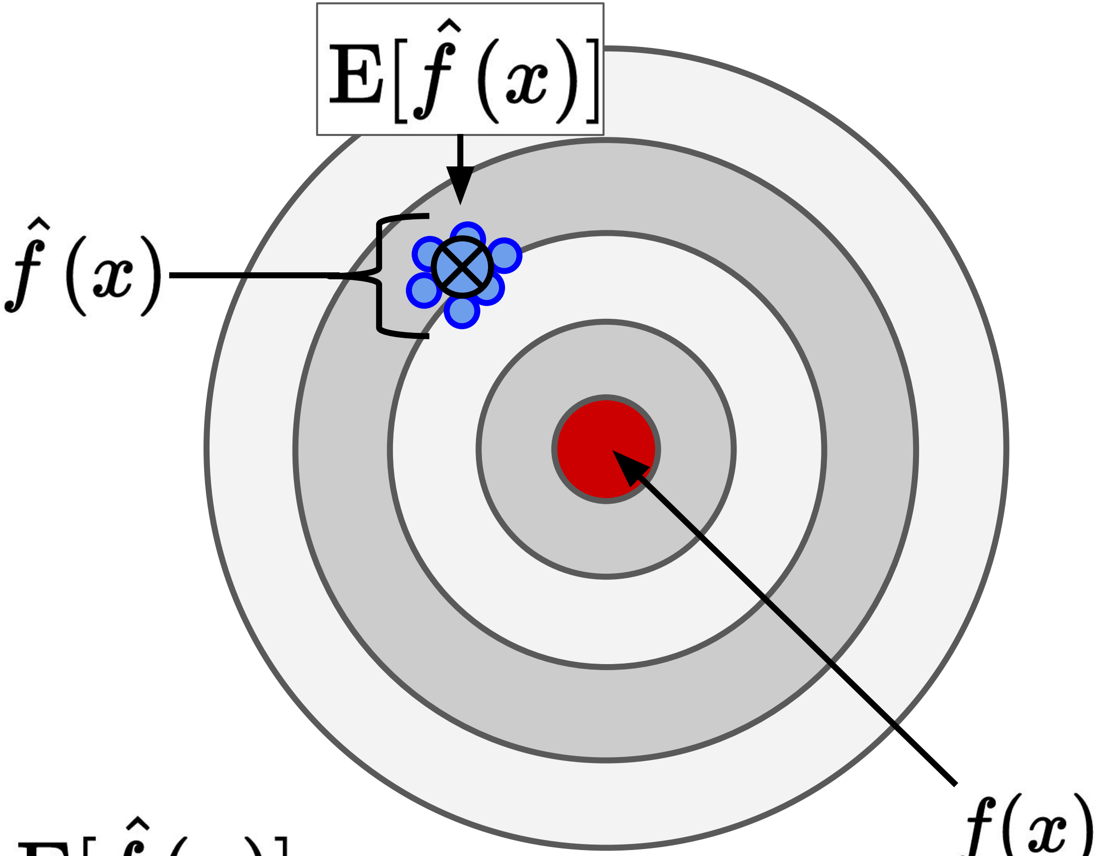
- What is the “true value”?
  the bullseye
- What are the “results”?
  where the arrows landed
- What’s the model?
  the archer

Adding some statistical notation:

true value: $f(x)$

results: $\hat{f}(x)$

the expected value of the results: $\mathbf{E}[\hat{f}(x)]$

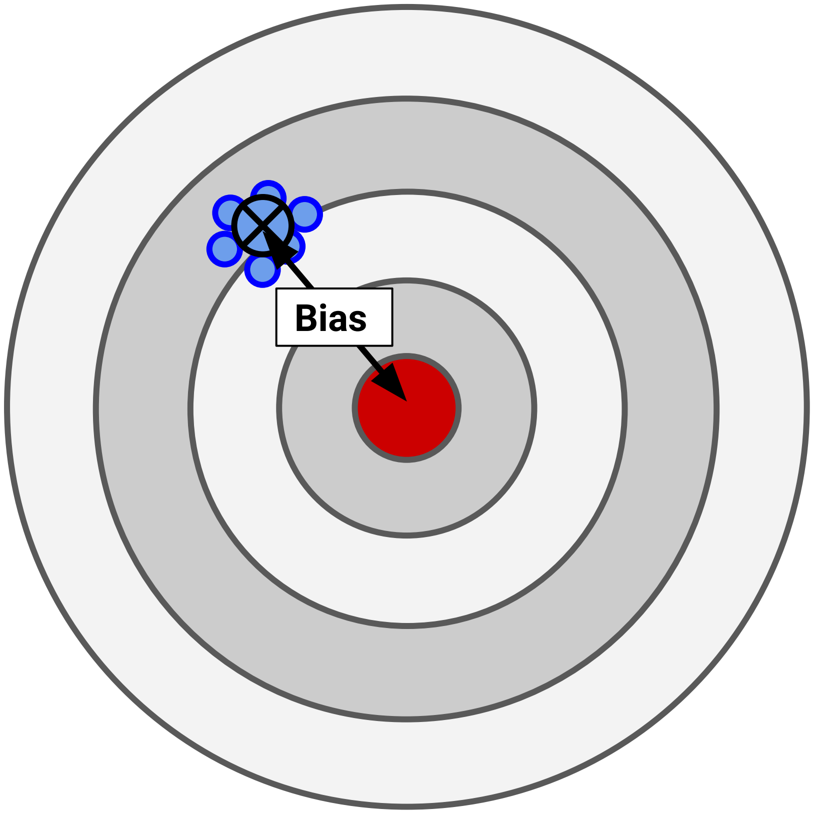


Bias and Variance Introduction

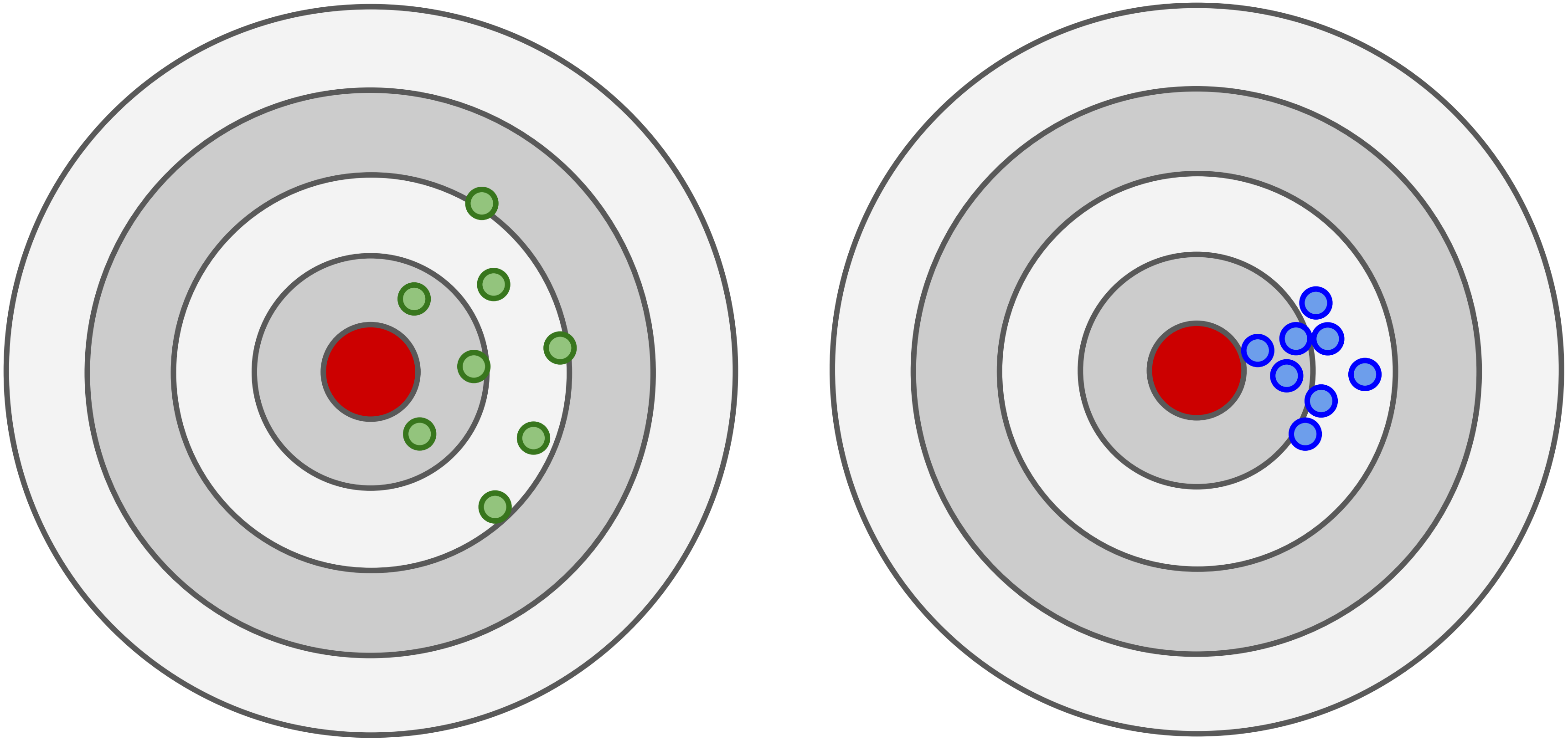
Bias:

The amount the **expected value** of the results differ from the true value.

$$\mathrm{Bias}\left[\hat{f}(x)\right] = \mathbf{E}\left[\hat{f}(x) - f(x)\right]$$

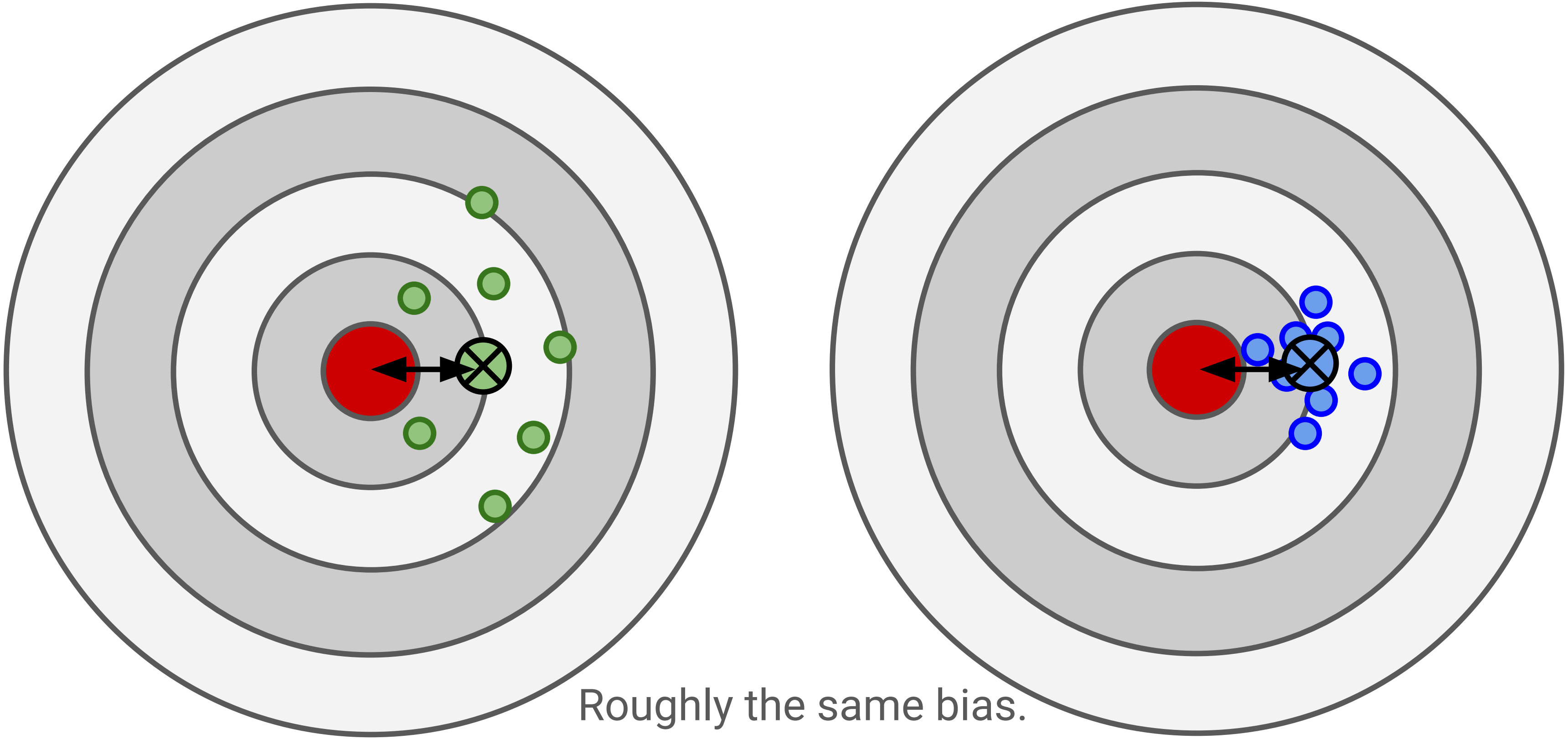


Bias and Variance Introduction



Compare the biases of these two archers (models)

Bias and Variance Introduction



Roughly the same bias.

Difference between expected value of result and true value are the same in both cases.

But what's not the same?

Bias and Variance Introduction

Bias:

The amount the expected value of the results differ from the true value.

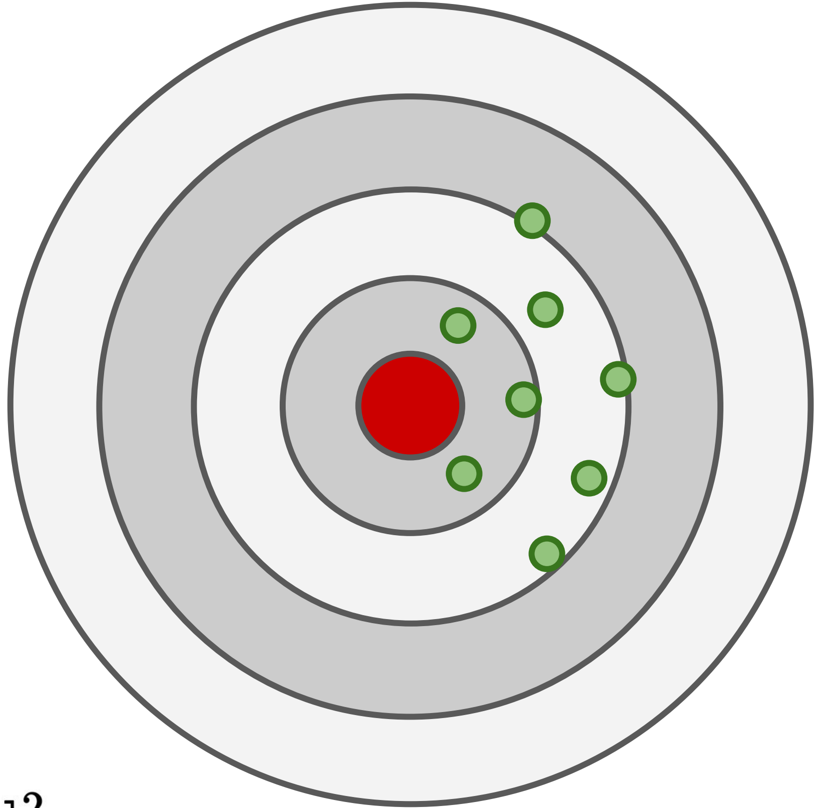
$$\mathrm{Bias}\left[\hat{f}(x)\right] = \mathbf{E}\left[\hat{f}(x) - f(x)\right]$$

Variance:

The expected value of the squared deviation of the results from the mean of the results.

$$\mathrm{Var}\left[\hat{f}(x)\right] = \mathbf{E}[\hat{f}(x)^2] - \mathbf{E}[\hat{f}(x)]^2$$

derivation

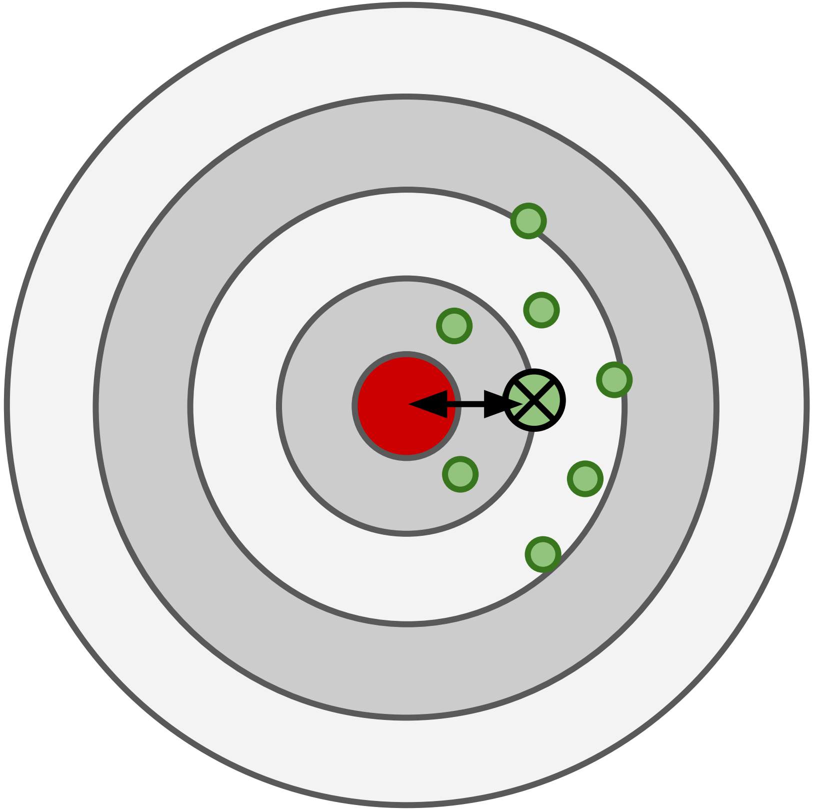


Bias and Variance Introduction

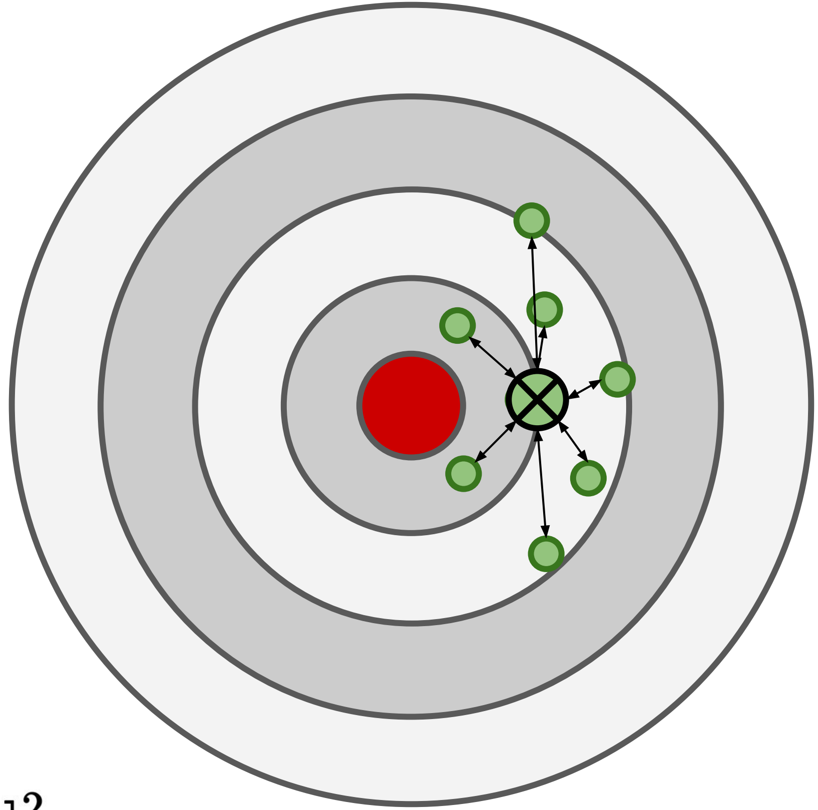
Bias:

The amount the expected value of the results differ from the true value.

$$\mathrm{Bias}\left[\hat{f}\,(x)\right] = \mathbf{E}\left[\hat{f}\,(x) - f(x)\right]$$



Bias and Variance Introduction



Variance:

The expected value of the squared deviation of the results from the mean of the results.*

$$\mathbf{Var}\left[\hat{f}\,(x)\right] = \mathbf{E}[\hat{f}\,(x)^2] - \mathbf{E}[\hat{f}\,(x)]^2$$

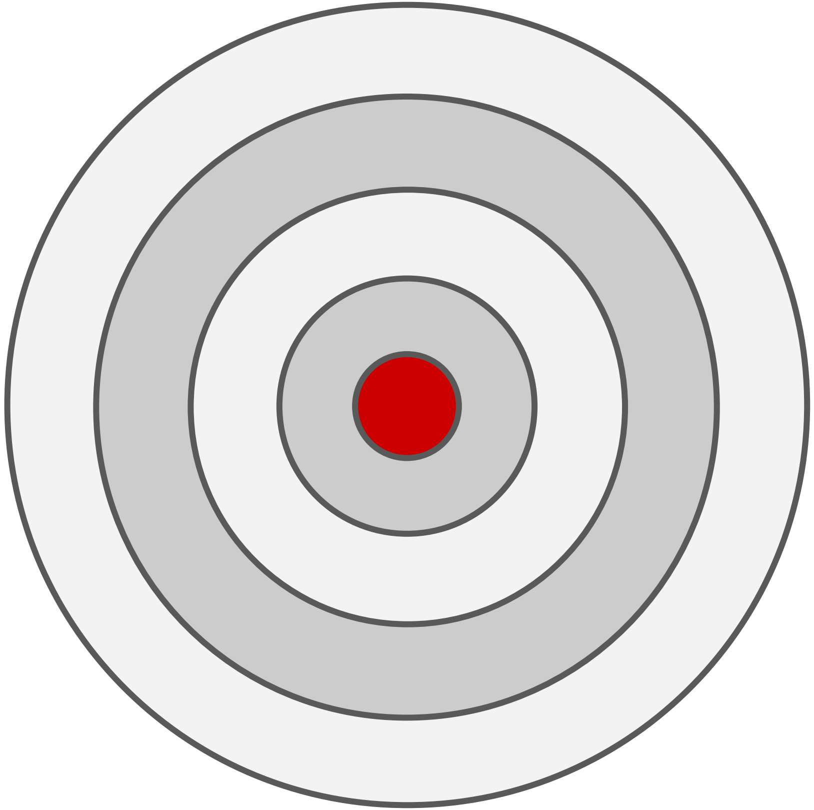
*no mention of true value!

Bias and Variance: Breakout

In groups of two at your tables, draw archery results illustrating the following 4 scenarios:

A) Low bias, low variance
B) Medium bias, high variance
C) High bias, low variance
D) Low bias, high variance

Who’s the best archer?



The Bias-Variance Trade-off

See jupyter notebook:

demo_bias_variance_tradeoff.ipynb

Verbalizing the Bias-Variance Trade-off

A low **bias** model accurately predicts the population true value.

A low **variance** model’s predictions don’t change much when fit on different samples from the population.

A trade-off often exists between bias and variance.
Some amount of model complexity is required to capture the true population signal, yet too much complexity yields a model highly sensitive to perturbations in sample data in the model fitting procedure.

So as bias decreases, variance often increases (and vice-versa).