Bias-Variance Tradeoff

Machine Learning

Bias and variance

Every learning algorithm requires assumptions about the hypothesis space.

Eg: “My hypothesis space is

- …linear”
- …decision trees with 5 nodes”
- …quadratic kernel”

Bias and variance

Every learning algorithm requires assumptions about the hypothesis space.

Eg: “My hypothesis space is

- …linear”
- …decision trees with 5 nodes”
- …quadratic kernel”

Bias is the true error (loss) of the *best* predictor in the hypothesis set

Bias and variance

Every learning algorithm requires assumptions about the hypothesis space.

Eg: “My hypothesis space is

- …linear”
- …decision trees with 5 nodes”
- …quadratic kernel”

Bias is the true error (loss) of the *best* predictor in the hypothesis set

- What will the bias be if the hypothesis set can not represent the target function? (high or low?)

Bias and variance

Every learning algorithm requires assumptions about the hypothesis space.

Eg: "My hypothesis space is

- …linear"
- …decision trees with 5 nodes"
- …quadratic kernel"

Bias is the true error (loss) of the *best* predictor in the hypothesis set

- What will the bias be if the hypothesis set can not represent the target function? (high or low?)
  - Bias will be non zero, possibly high
- Underfitting: When bias is high

Bias and variance

- The performance of a classifier is dependent on the specific training set we have
  - Perhaps the model will change if we slightly change the training set

Bias and variance

- The performance of a classifier is dependent on the specific training set we have
  - Perhaps the model will change if we slightly change the training set

- Variance: Describes how much the best classifier depends on the training set

Bias and variance

- The performance of a classifier is dependent on the specific training set we have
  - Perhaps the model will change if we slightly change the training set

- Variance: Describes how much the best classifier depends on the training set

- Overfitting: High variance

Bias and variance

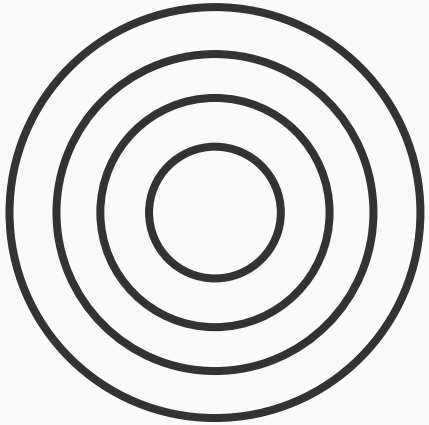
- The performance of a classifier is dependent on the specific training set we have
  - Perhaps the model will change if we slightly change the training set
- Variance: Describes how much the best classifier depends on the training set
- Overfitting: High variance
- Variance
  - Increases when the classifiers become more complex
  - Decreases with larger training sets

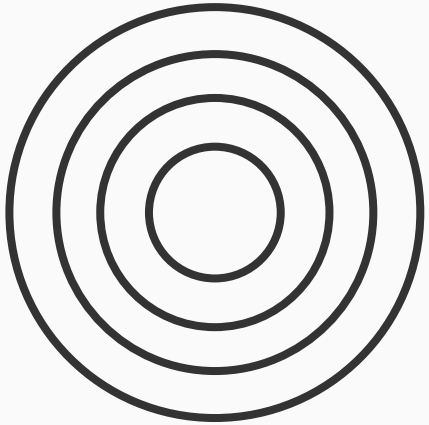
Let’s play darts

Suppose the true concept is the center

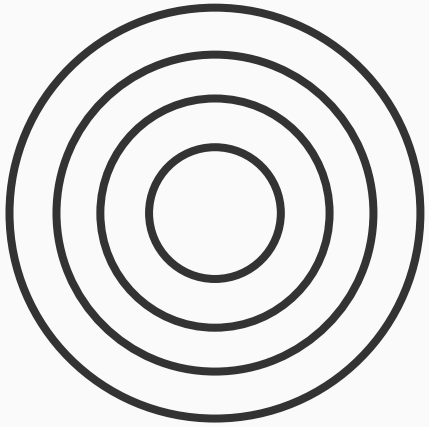
Each dot is a model that is learned from a different dataset

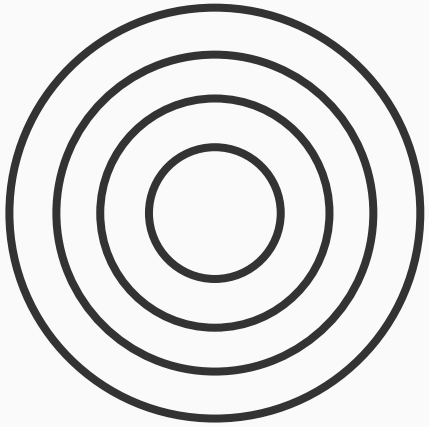
High bias





Low bias





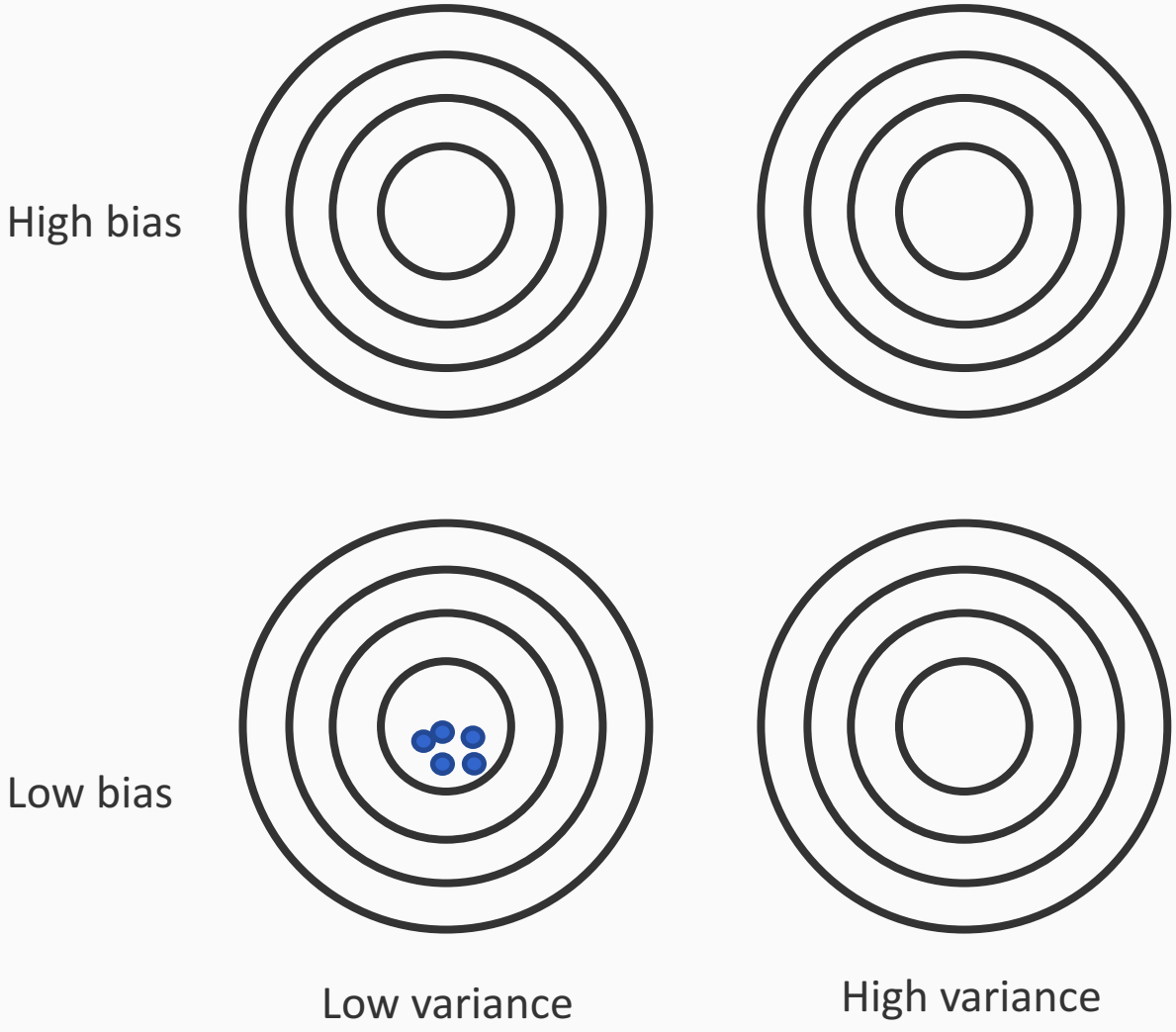
Low variance

High variance

Let's play darts

Suppose the true concept is the center

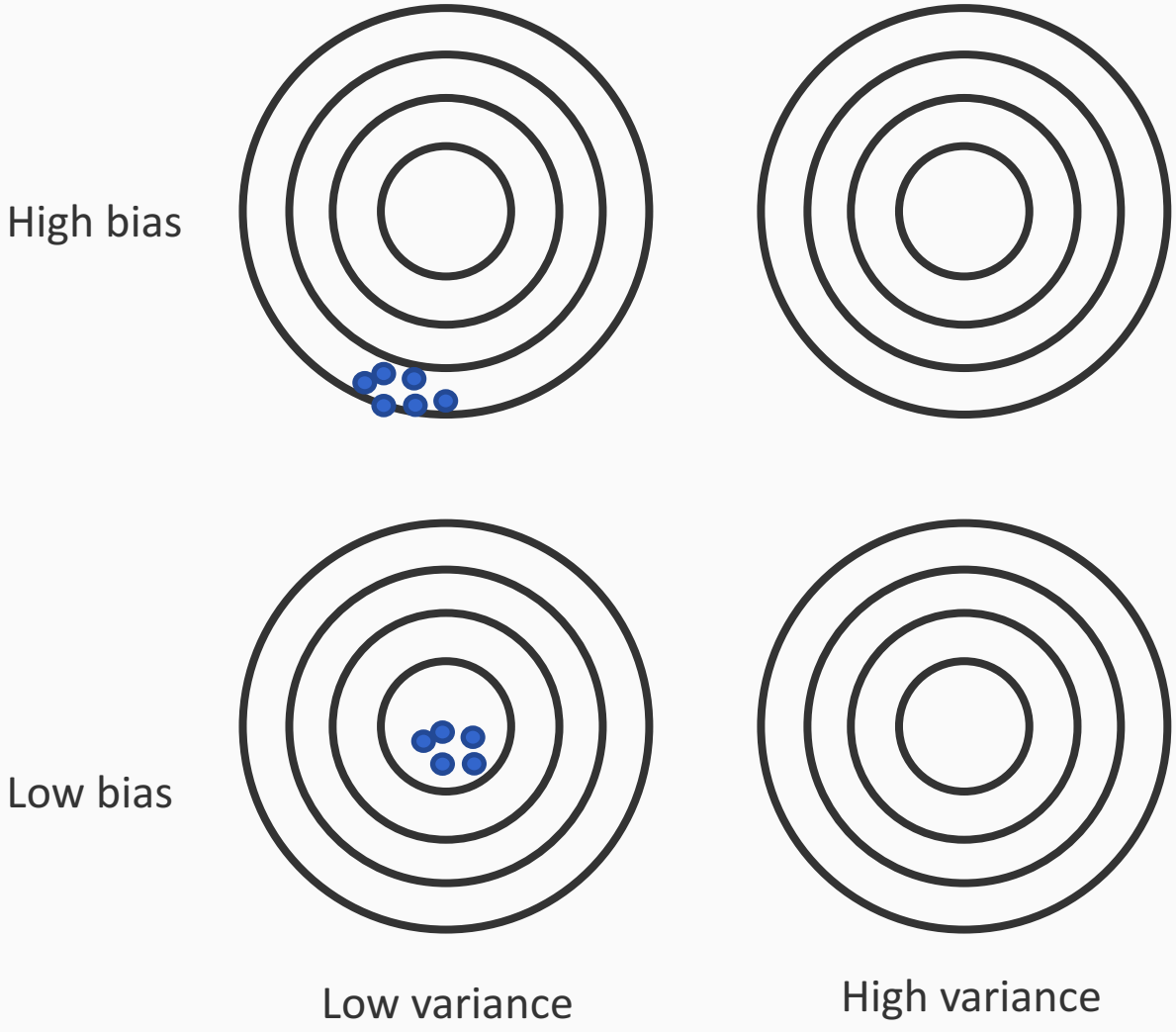
Each dot is a model that is learned from a different dataset



Let's play darts

Suppose the true concept is the center

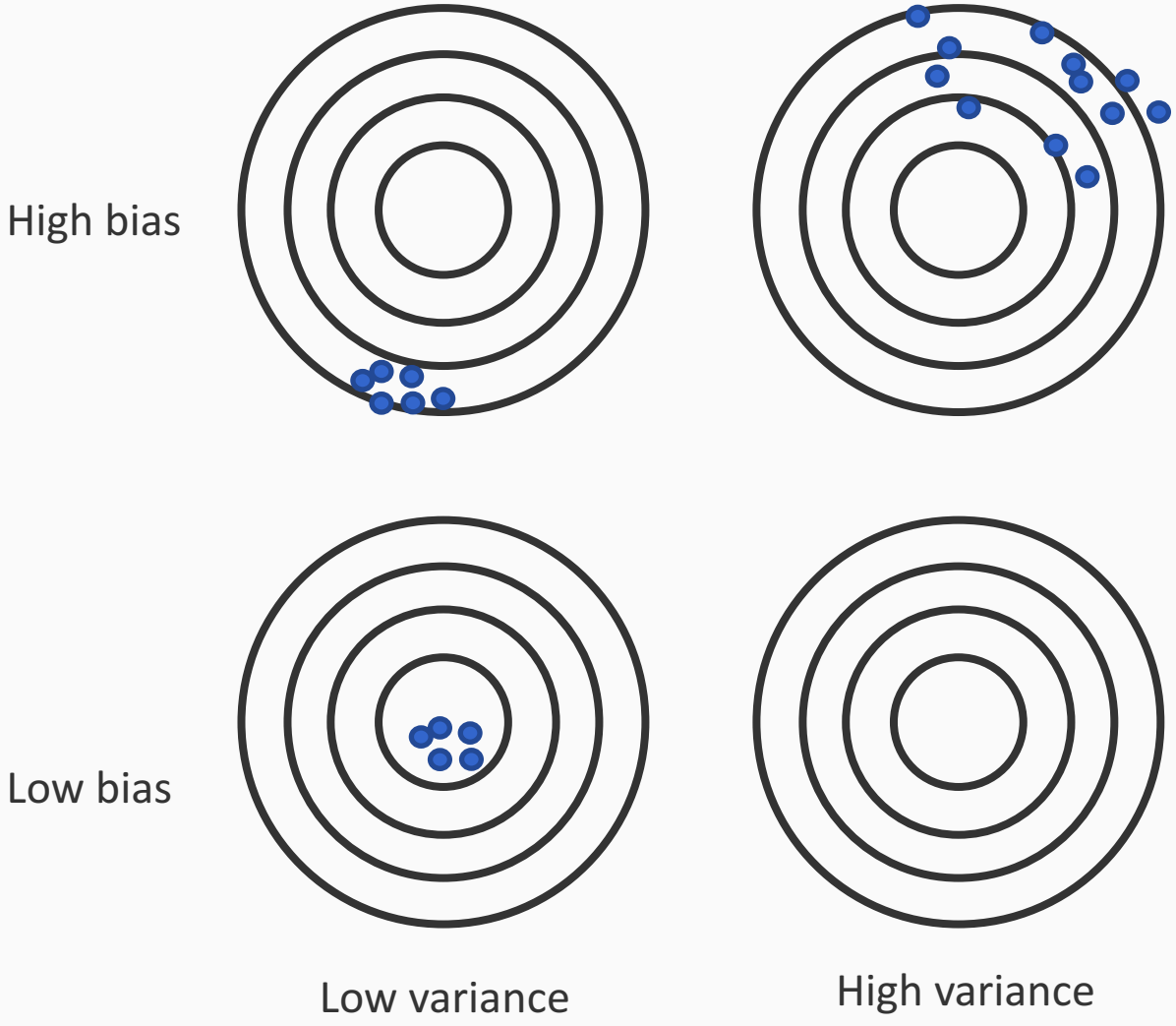
Each dot is a model that is learned from a different dataset



Let's play darts

Suppose the true concept is the center

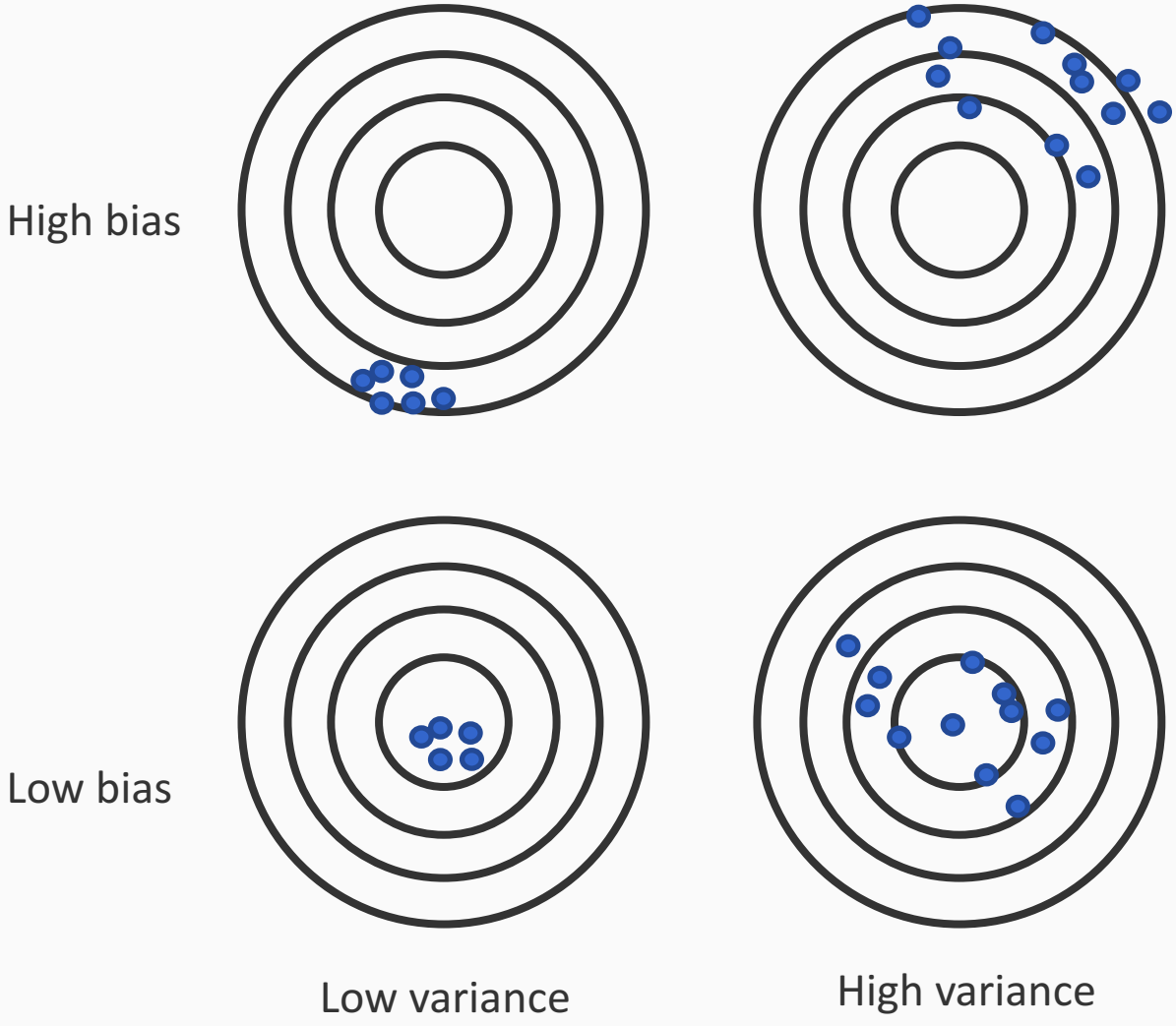
Each dot is a model that is learned from a different dataset



Let’s play darts

Suppose the true concept is the center

Each dot is a model that is learned from a different dataset



Bias variance tradeoffs

- Error = bias + variance (+ noise)
- High bias $\Rightarrow$ both training and test error can be high
  - Arises when the classifier can not represent the data
- High variance $\Rightarrow$ training error can be low, but test error will be high
  - Arises when the learner overfits the training set

Bias variance tradeoff has been studied extensively in the context of regression
Generalized to classification (Pedro Domingos, 2000)

Managing of bias and variance

- Ensemble methods reduce variance
  - Multiple classifiers are combined
  - Eg: Bagging, boosting
- Decision trees of a given depth
  - Increasing depth decreases bias, increases variance
- SVMs
  - Higher degree polynomial kernels decreases bias, increases variance
  - Stronger regularization increases bias, decreases variance
- K nearest neighbors (we will see this at the end of the semester)
  - Increasing k generally increases bias, reduces variance

Summary

- Bias and Variance
  - Rich exploration in statistics
  - Provides a different view of learning criteria