Overview

- Types of statistical learning problems

- Why learning is difficult?

- Bias variance tradeoff

- An example: kNN *vs* Linear Regression (in a separate pdf file)
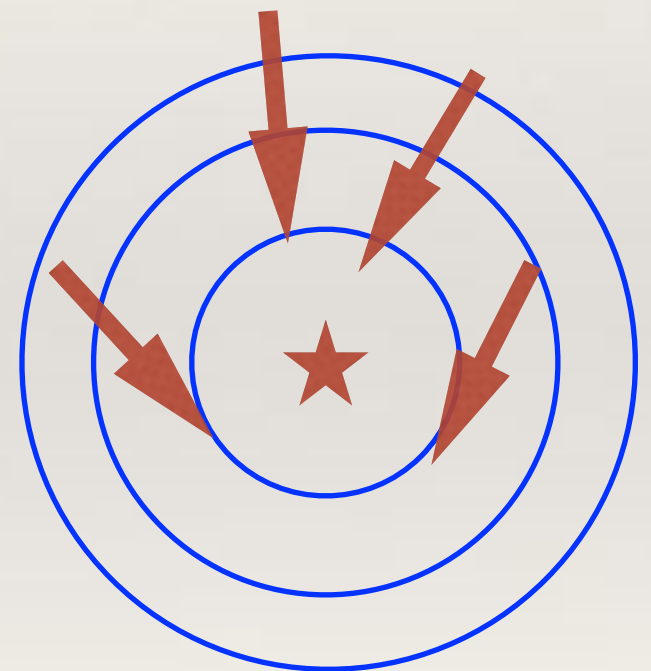
- Not all about prediction

Bias Variance Tradeoff

Goal of ML: Minimize *generalization error* (i.e., error on unseen future datasets), not training error.

Source of errors:
- ❖ Bias
- ❖ Variance
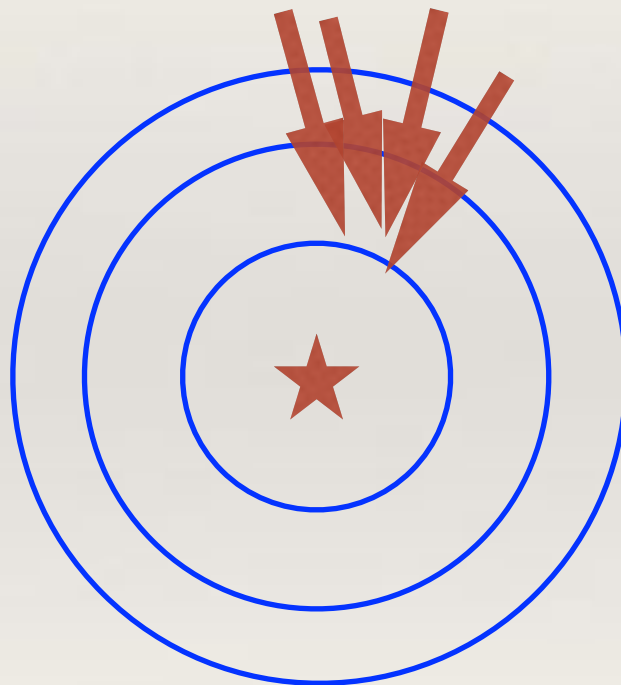
Bias Variance Tradeoff

Goal of ML: Minimize *generalization error* (i.e., error on unseen future datasets), not training error.

Source of errors:
- ❖ Bias
- ❖ Variance

Truth ★

$\hat{f}_n$
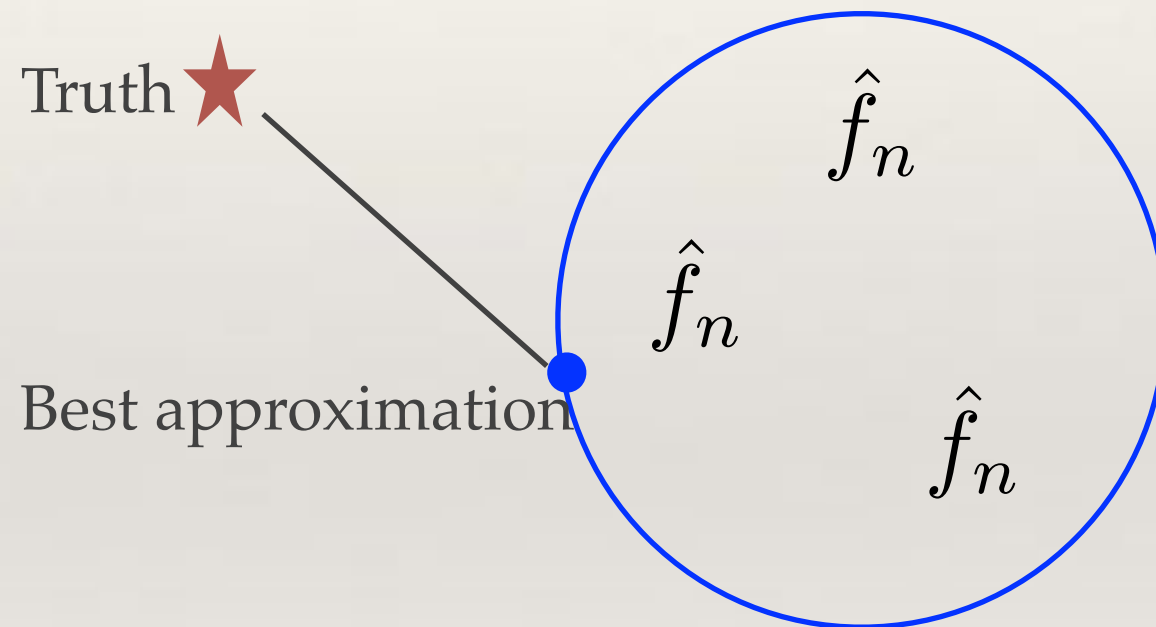
$\hat{f}_n$

Best approximation

$\hat{f}_n$

Bias Variance Tradeoff

Goal of ML: Minimize *generalization error* (i.e., error on unseen future datasets), not training error.

Source of errors:
- ❖ Bias
- ❖ Variance

Truth ★

$\hat{f}_n$

$\hat{f}_n$
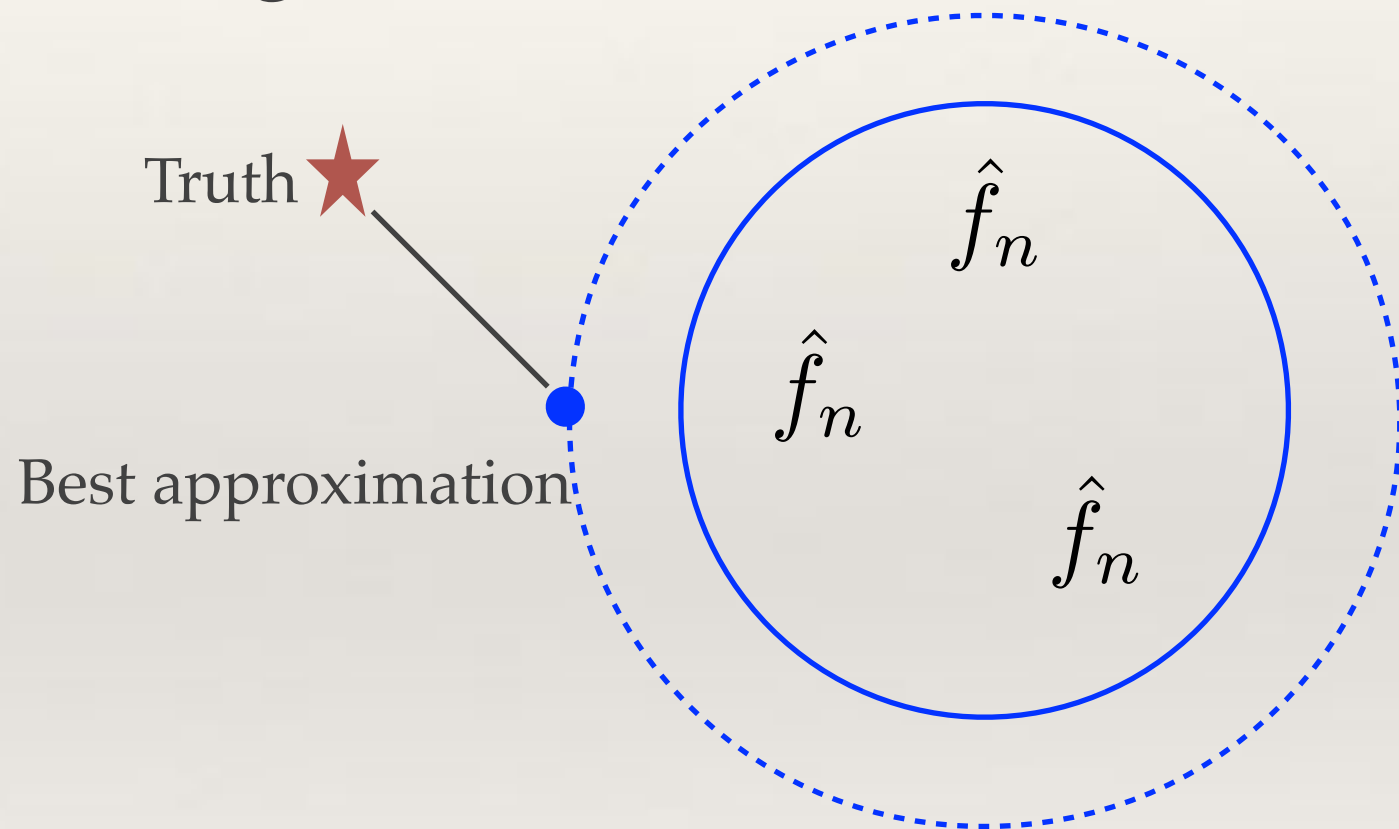
$\hat{f}_n$

Best approximation

Bias Variance Tradeoff

Goal of ML: Minimize *generalization error* (i.e., error on unseen future datasets), not training error.

Source of errors:
- ❖ Bias
- ❖ Variance

Truth ★

$\hat{f}_n$

$\hat{f}_n$

$\hat{f}_n$

$\hat{f}_n$

Best approximation

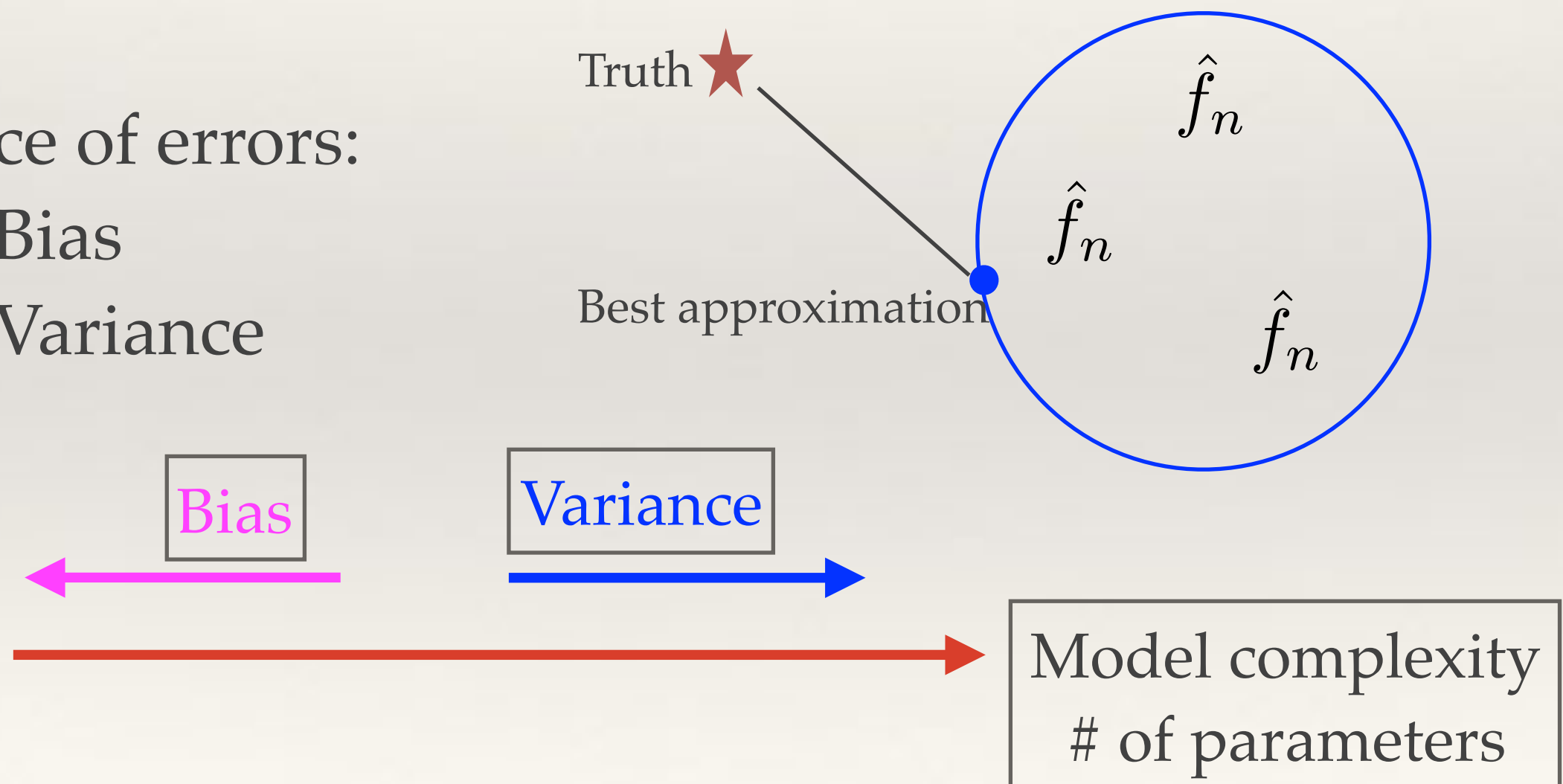| Bias |

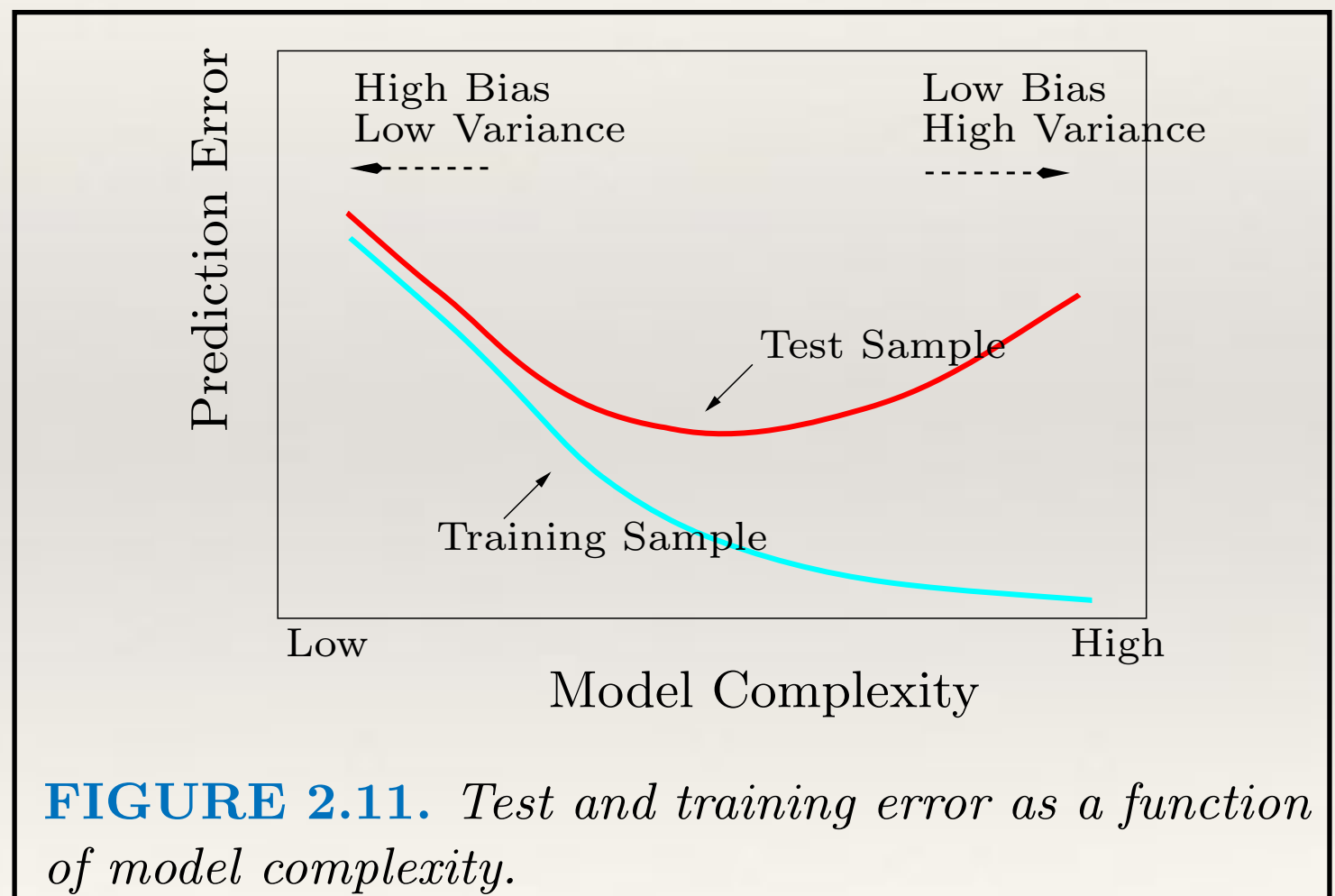| Variance |

Model complexity
# of parameters

Bias Variance Tradeoff

Goal of ML: Minimize *generalization error* (i.e., error on unseen future datasets), not training error.

Source of errors:
- ❖ Bias
- ❖ Variance

FIGURE 2.11. *Test and training error as a function of model complexity.*
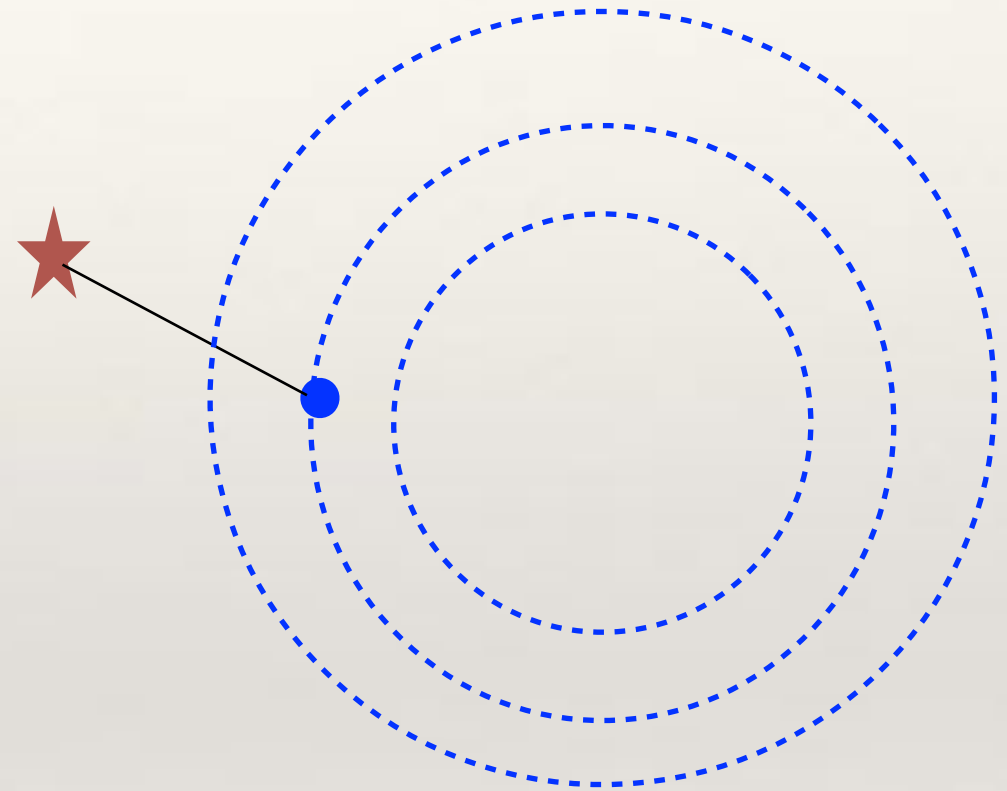
What'll be Covered in Stat542

❖ Flexible modeling techniques to reduce bias

❖ Useful strategies to achieve the tradeoff between bias and variance
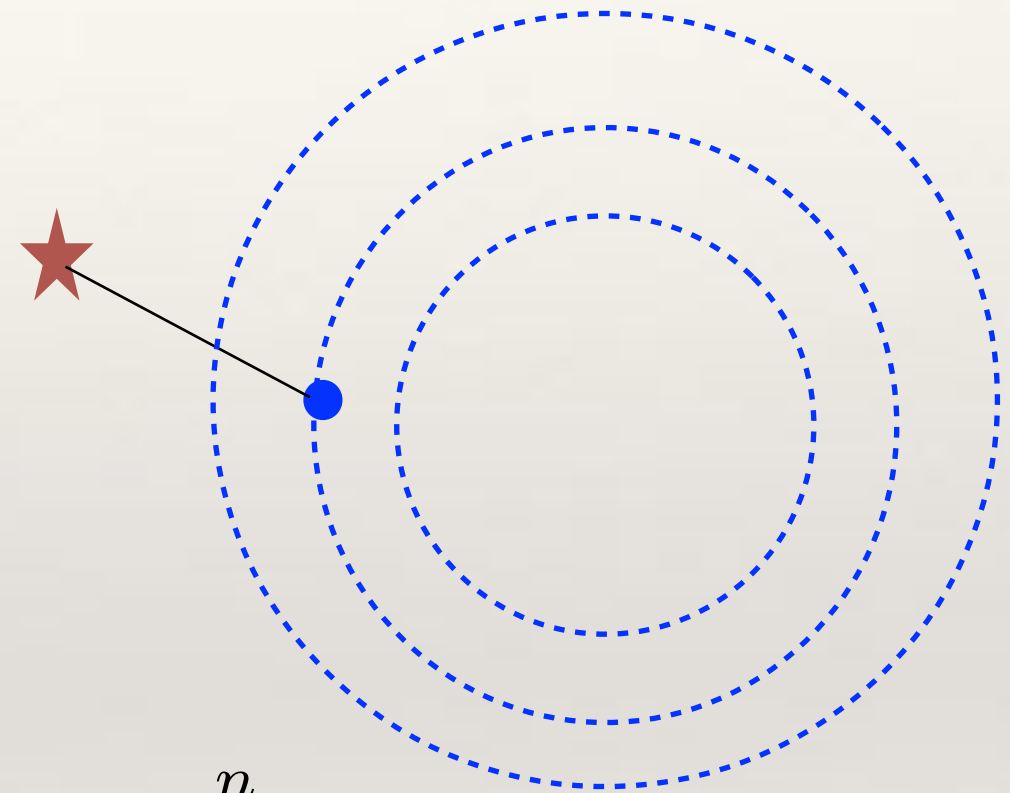
Two Successful Strategies

❖ <u>Regularization</u>: Restrict the parameters to a low-dimensional space, which is *adaptively* determined by the data.

Two Successful Strategies

❖ <u>Regularization</u>: Restrict the parameters to a low-dimensional space, which is *adaptively* determined by the data.

$$\min_{w} \left[\sum_{i=1}^{n} \text{loss}(y_i, f(x_i; w) + \lambda |w| \right]$$

LASSO

Two Successful Strategies

❖ <u>Regularization</u>: Restrict the parameters to a low-dimensional space, which is *adaptively* determined by the data.
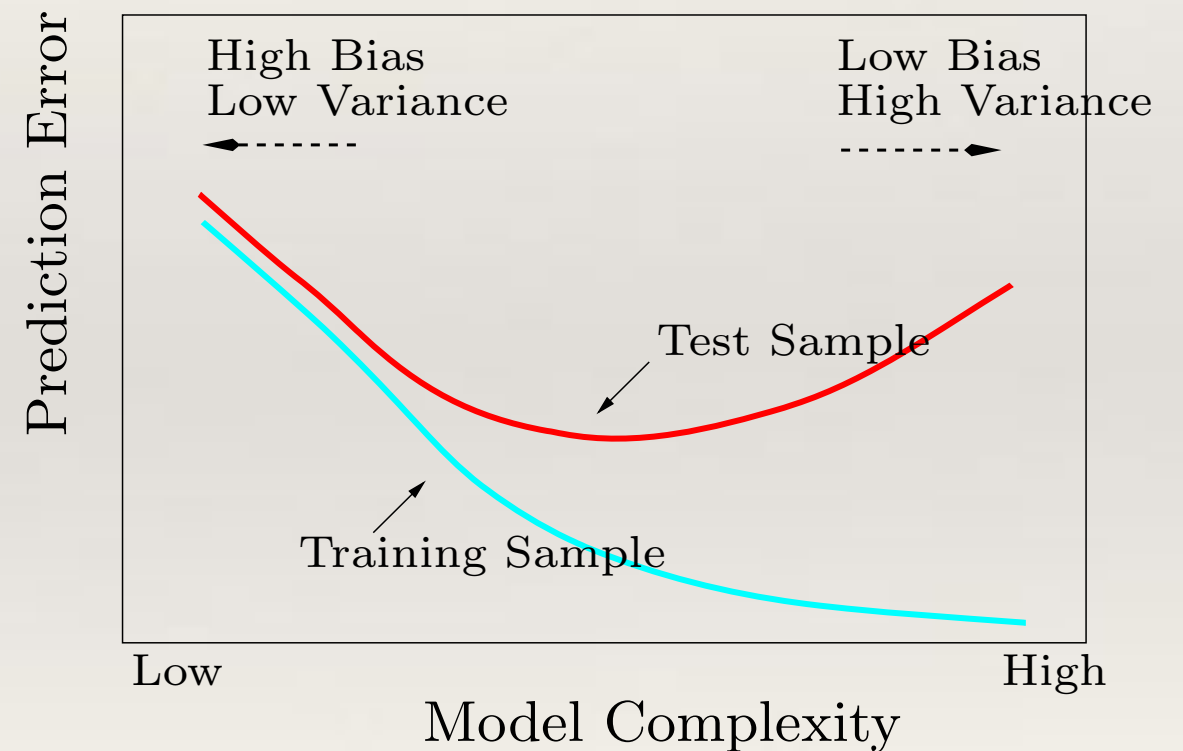
❖ <u>Ensemble:</u> Average many low-bias high-variance models; averaging reduces variance.

FIGURE 2.11. *Test and training error as a function of model complexity.*

Overview

❖ Types of statistical learning problems

❖ Why learning is difficult?

❖ Bias variance tradeoff

❖ An example: kNN *vs* Linear Regression

❖ Not all about prediction

Not All About Prediction

- ❖ Although the focus of this course is prediction, statistical learning ≠ prediction

- ❖ Exploration *vs.* Prediction

- ❖ Data product *vs.* decision making

- ❖ Make your model to generate actionable insights