
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

Problems (I)

- ❖ Project 1 (Ames Housing Data): Predict the sale price of a house given its features.
- ❖ Project 2 (Sales Forecasting): Provide sales forecasting for Walmart for each department in each store based on historical data.

Y : Target

X : Feature / Covariates

Problems (II)

- ❖ Project 3 (Lending Club): Determine the chance that a borrower will miss a payment next month given various characteristics of the borrower and the loan.
- ❖ Project 4 (Sentiment Analysis): Determine whether a movie review is positive or negative.

Y : Target

X : Feature / Covariates

Problems (III)

- ❖ Based on the recent real estate transactions at Ames, Iowa, can we identify any home buying / selling trends? Further, can we identify distinctive groups of buyers?
- ❖ Based on the transaction data at Walmart, can we recommend any marketing strategies to Walmart?

[Association Rule](#) (chap 14.2 of ESL)

[Market Segmentation](#) (cluster customers)

Problems (III)

- ❖ Based on the recent real estate transactions at Ames, Iowa, can we identify any home buying / selling trends? Further, can we identify distinctive groups of buyers?
- ❖ Based on the transaction data at Walmart, can we recommend any marketing strategies to Walmart?

~~Y: Target~~

X: Feature / Covariates

Types of Statistical Learning Problems

- ❖ Supervised Learning

- ❖ Regression: response is a number

- ❖ Classification: response is a label (binary or multi-class)

Semi-supervised Learning

Recommender System

- ❖ Unsupervised Learning: identify latent structures in the data, e.g., clustering, association rule, HMM, etc.

How does Supervised Learning Work?

Data: $\{x_i, y_i\}$
 $i=1, 2, \dots, n$

Goal: learn w
 $y \leftarrow f(x; w)$

Minimize an objective function

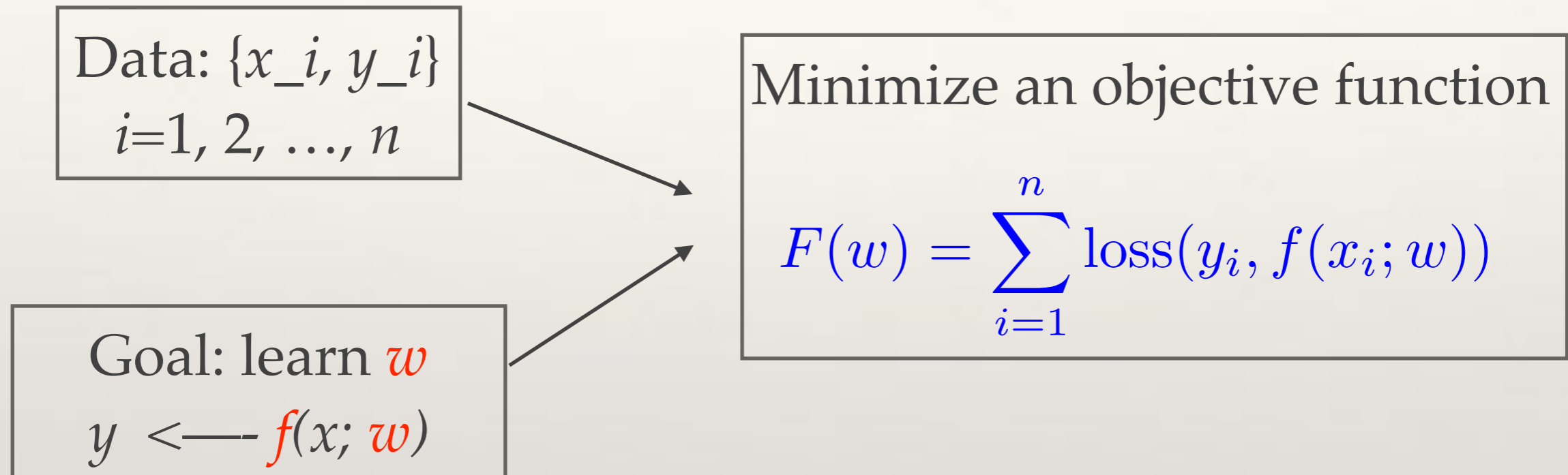
$$F(w) = \sum_{i=1}^n \text{loss}(y_i, f(x_i; w))$$

$$\text{loss} = [y_i - f(x_i; w)]^2$$

$$\text{loss} = \mathbf{1} \{y_i \neq f(x_i; w)\}$$

$$\text{loss} = \begin{cases} -\log f(x_i; w) & \text{if } y_i = 1 \\ -\log(1 - f(x_i; w)) & \text{if } y_i = 0 \end{cases}$$

How does Supervised Learning Work?



1. The minimizer w^* may be in closed form.
2. Try optimization algorithms that can guarantee to converge to the global minimizer.
3. In the worst case, try *gradient descent*.

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

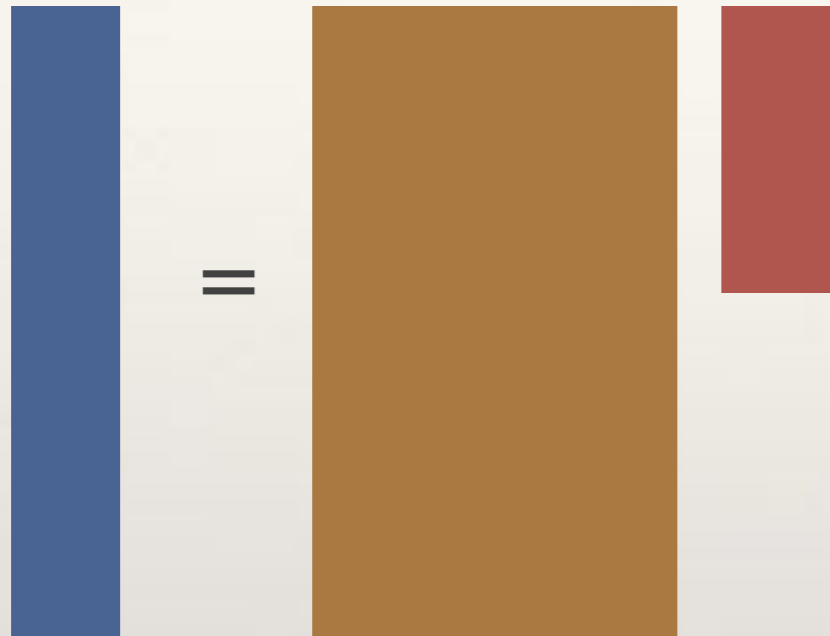
Challenges

- ❖ Training error underestimates test / generalization error.
- ❖ **Overfitting**: perform well on the training data but not on the future (unseen) data.
- ❖ p denotes the number of parameters the regression / classification function f has, i.e., the number of parameters we need to learn from the data.
- ❖ The gap between the two errors (training *vs.* test) gets large when p is large.

Curse of Dimensionality

- ❖ Curse of Dimensionality in Classification:
 - ❖ **1NN** (one-nearest-neighbor) predicts perfectly on the training data
 - ❖ Illustration on how dimensionality changes the performance of **linear classifiers**: <http://www.visiondummy.com/2014/04/curse-dimensionality-affect-classification/>

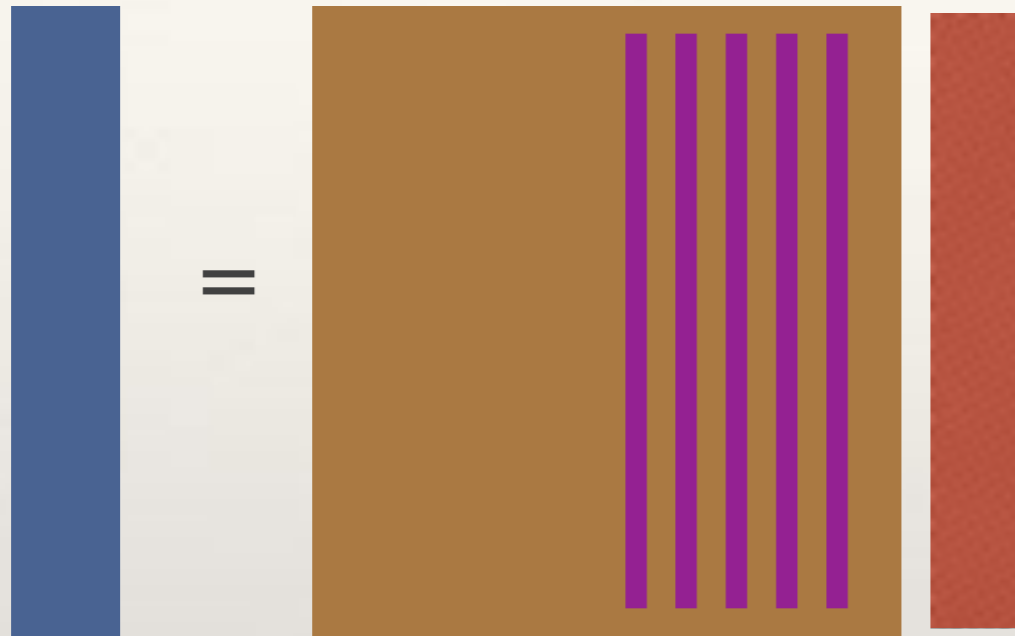
Curse of Dimensionality



$$y_{n \times 1} = X_{n \times p} w_{p \times 1}$$

Curse of Dimensionality in Regression

Curse of Dimensionality



$$y_{n \times 1} = X_{n \times n} w_{n \times 1}$$

n equations and n parameters
Perfect fit on the training data!

Overview

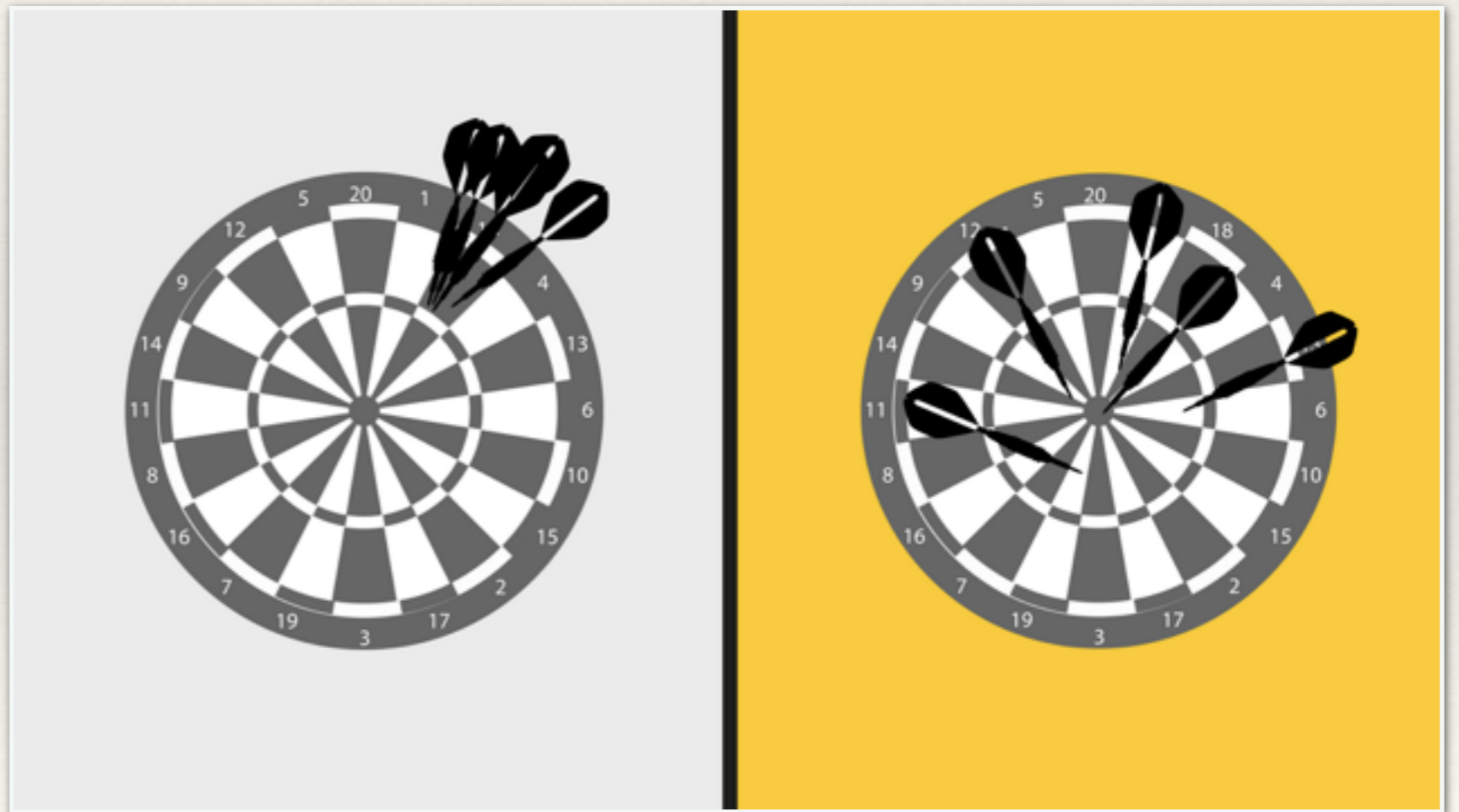
- ❖ Types of statistical learning problems
- ❖ Why learning is difficult?
- ❖ Bias variance tradeoff
- ❖ An example: kNN *vs* Linear Regression (in a separate pdf file)
- ❖ Not all about prediction

Bias Variance Tradeoff

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

Source of errors:

- ❖ Bias
- ❖ Variance

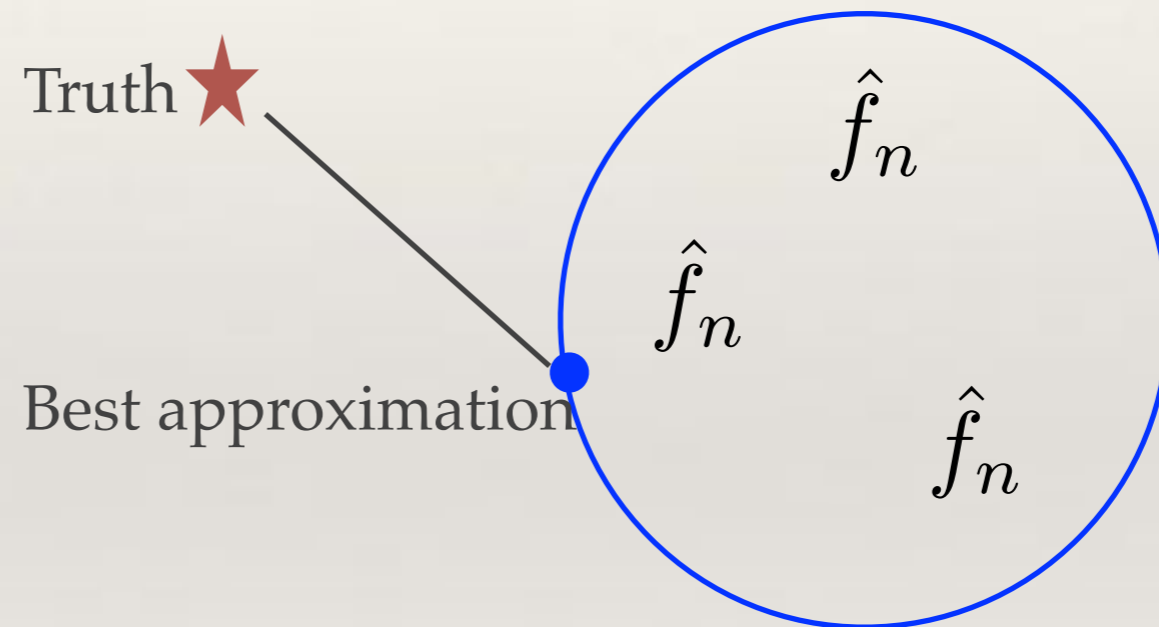


Bias Variance Tradeoff

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

Source of errors:

- ❖ Bias
- ❖ Variance

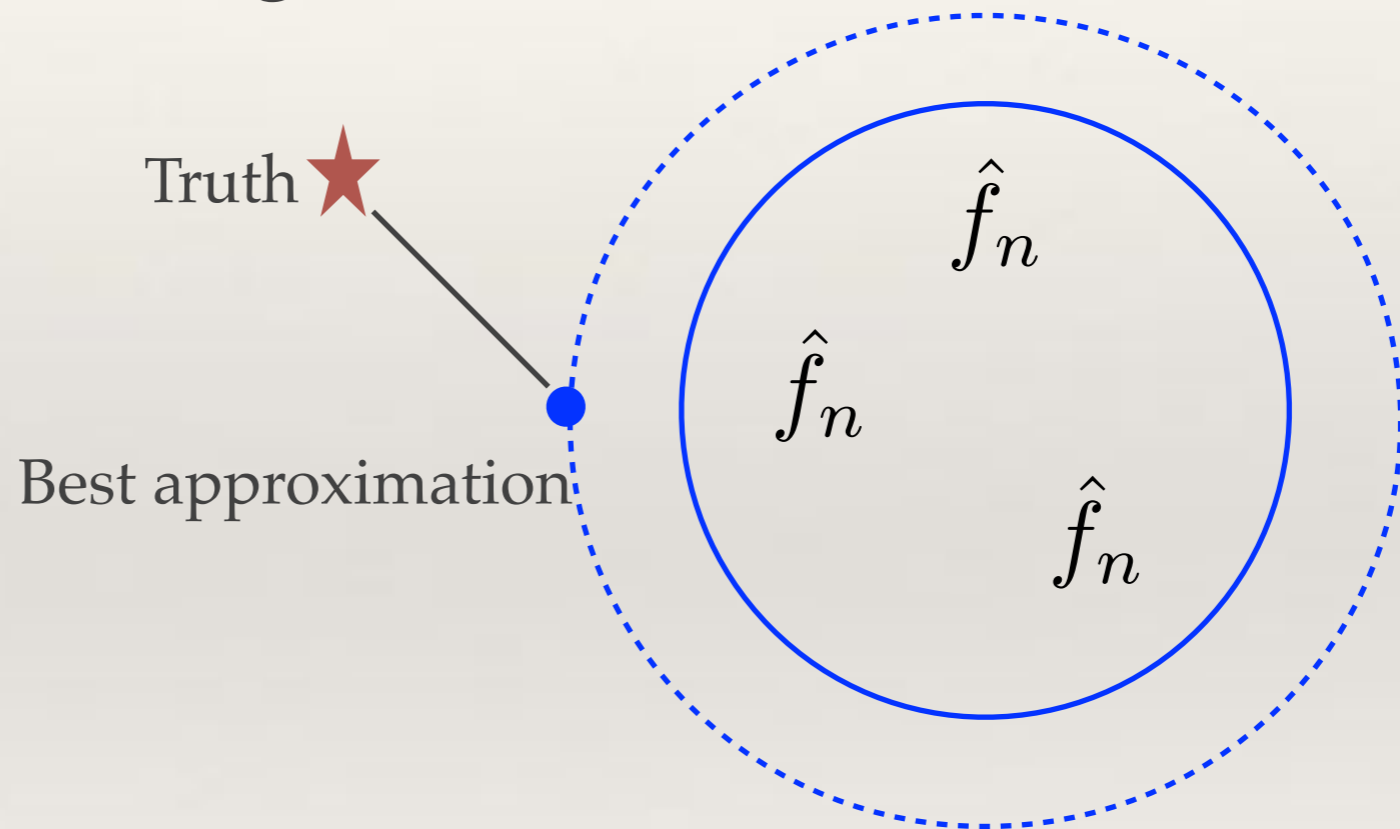


Bias Variance Tradeoff

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

Source of errors:

- ❖ Bias
- ❖ Variance

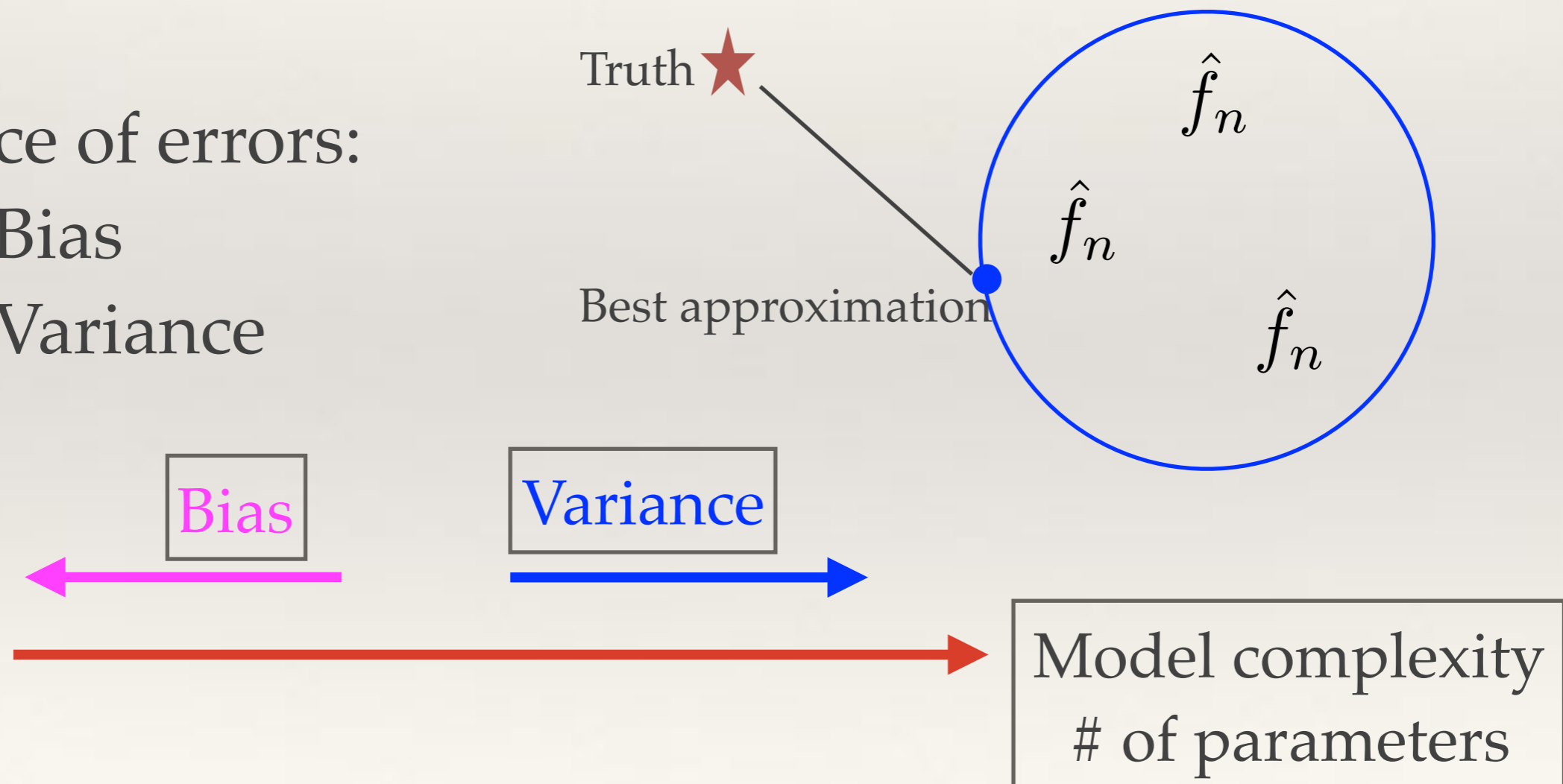


Bias Variance Tradeoff

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

Source of errors:

- ❖ Bias
- ❖ Variance



Bias Variance Tradeoff

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

Source of errors:

- ❖ Bias
- ❖ Variance

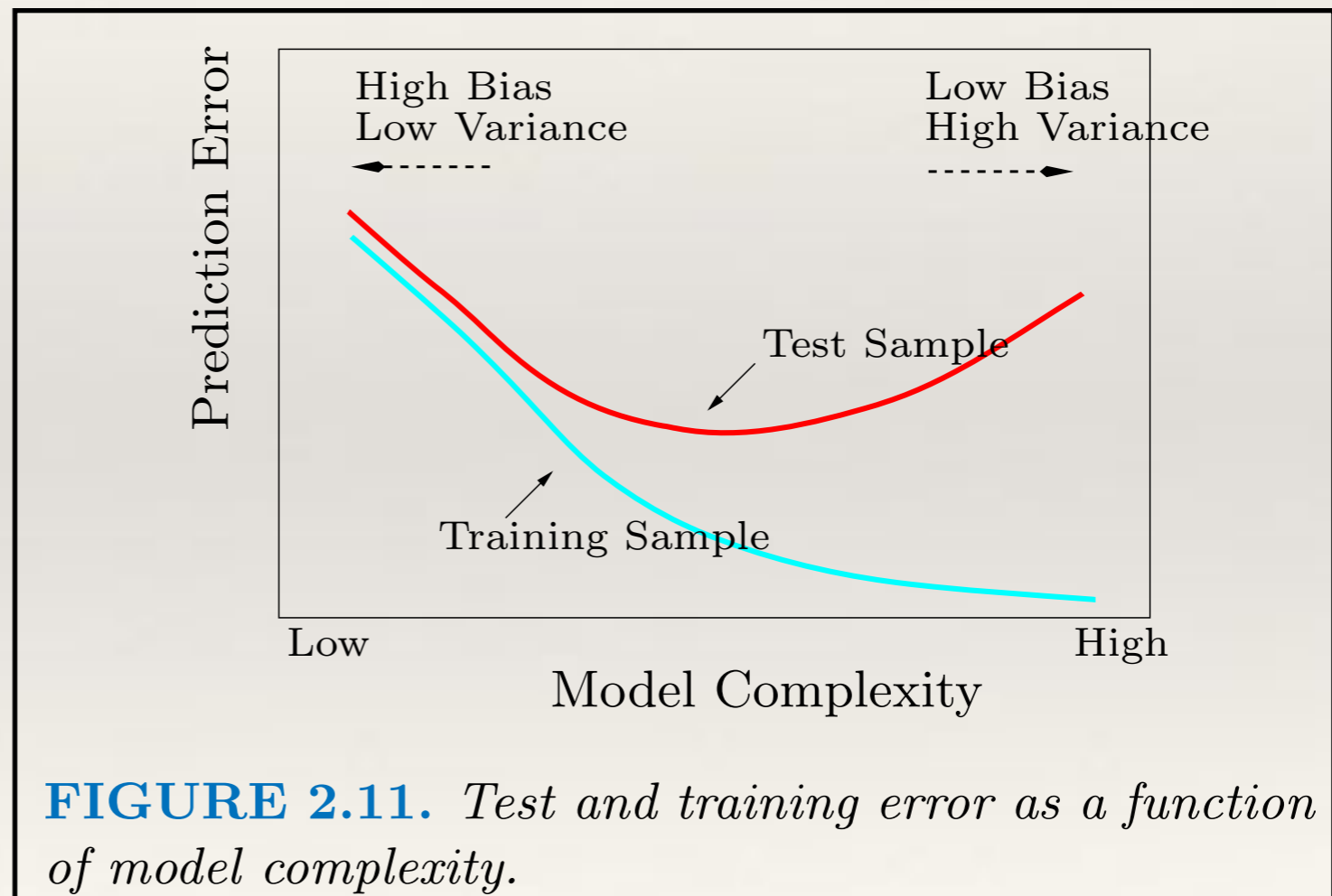


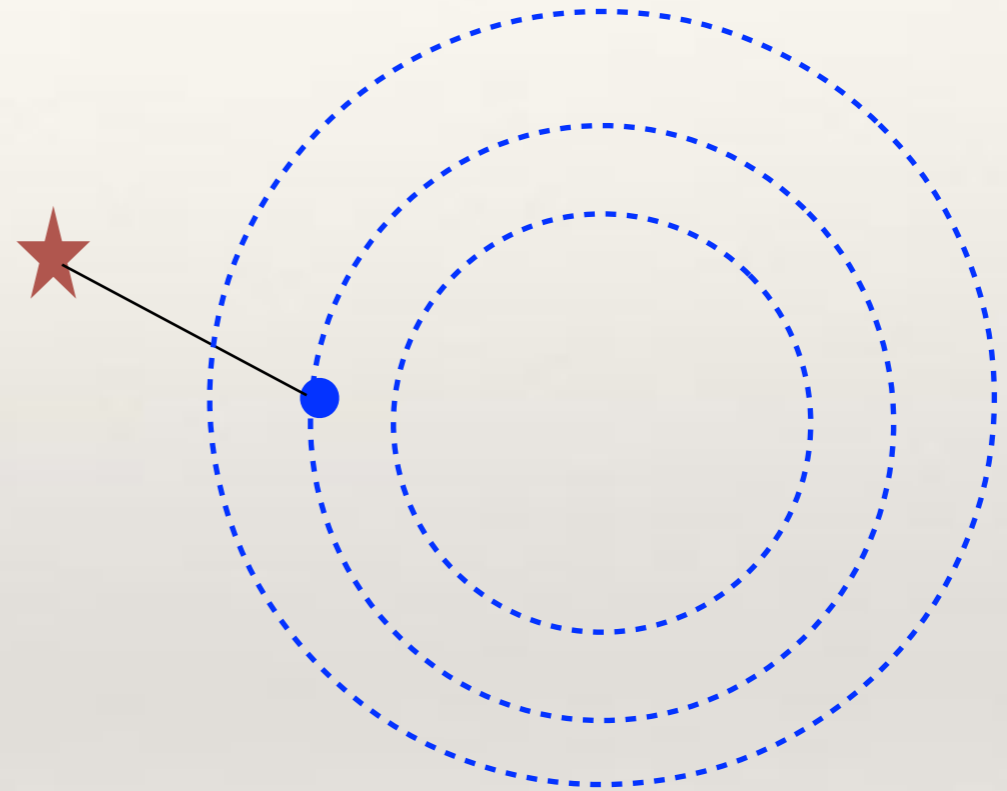
FIGURE 2.11. Test and training error as a function of model complexity.

What'll be Covered in Stat542

- ❖ Flexible modeling techniques to reduce bias
- ❖ Useful strategies to achieve the tradeoff between bias and variance

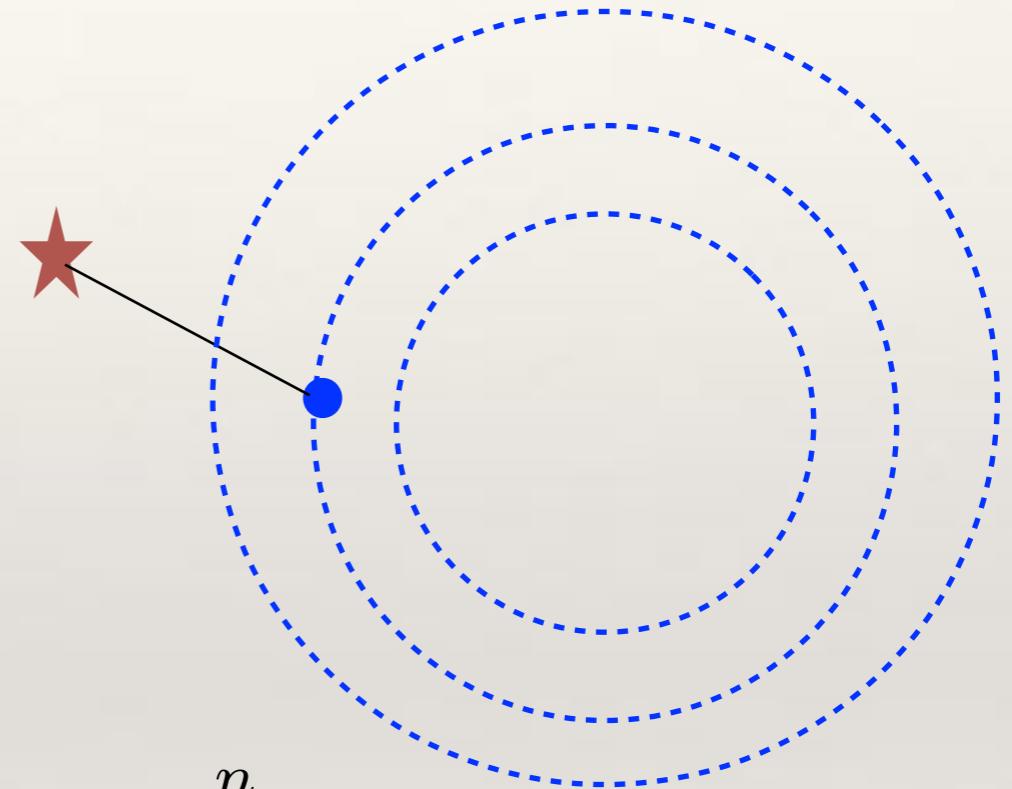
Two Successful Strategies

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



Two Successful Strategies

- ❖ Regularization: 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

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

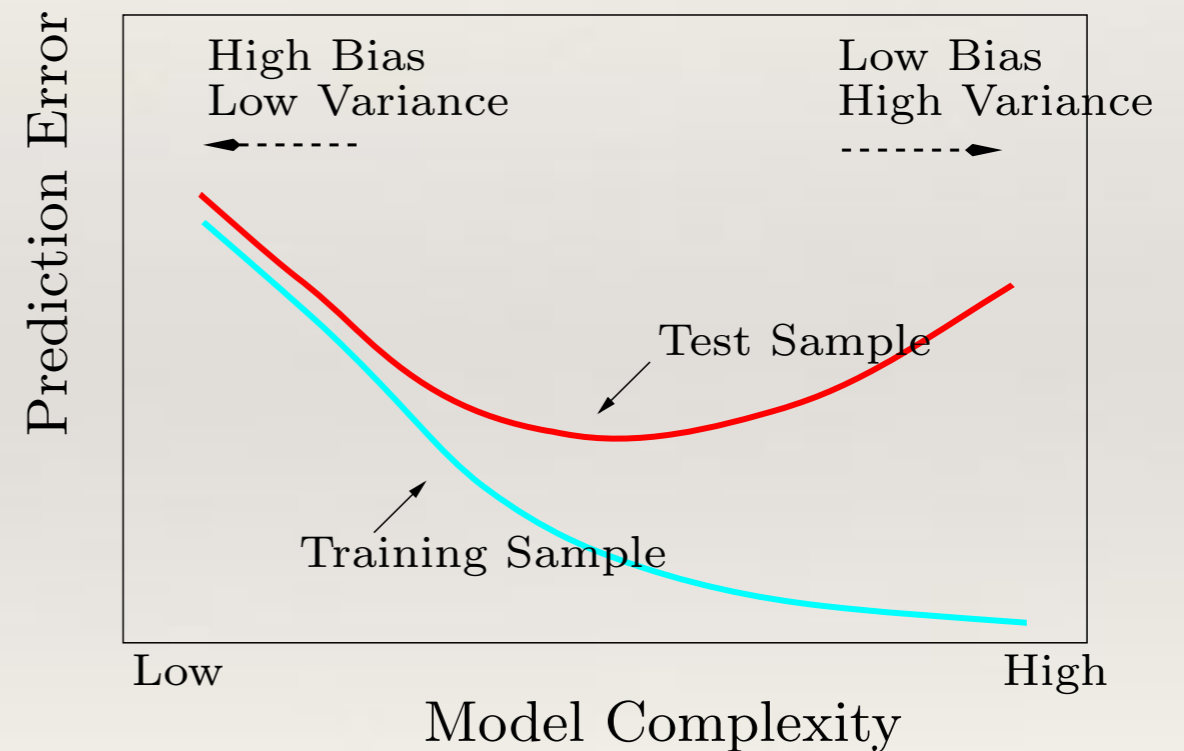


FIGURE 2.11. Test and training error as a function of model complexity.

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 \neq prediction
- ❖ Exploration *vs.* Prediction
- ❖ Data product *vs.* decision making
- ❖ Make your model to generate actionable insights