Introduction
DS351

Main principle

- Predictors $X = (X_1, X_2, \ldots, X_p)$
- Response Y

Main principle

- Predictors $X = (X_1, X_2, \ldots, X_p)$
- Response Y

Assumption: There's some function f and error ϵ such that

$$Y = f(X) + \epsilon$$

Here, Y and ϵ are **random variables**.

- Predictors $X = (X_1, X_2, \ldots, X_p)$
- Response Y

Let $\widehat{Y} = \hat{f}(X)$ where $\hat{f}$ is an estimate of f.

- Predictors $X = (X_1, X_2, \dots, X_p)$
- Response Y

Let $\widehat{Y} = \hat{f}(X)$ where $\hat{f}$ is an estimate of f.
Performance of $\hat{f}$ is measured by

$$\begin{aligned}\mathbf{E}(Y - \widehat{Y})^2 &= \mathbf{E}[f(X) + \epsilon - \hat{f}(X)]^2 \\ &= [f(X) - \hat{f}(X)]^2 + \mathrm{Var}(\epsilon).\end{aligned}$$

Our goals

- Minimize the reducible error.

Our goals

- Minimize the reducible error.
- Find relevant predictors.

Our goals

- Minimize the reducible error.
- Find relevant predictors.
- Find the relationship between X_i and Y e.g. would increasing the value of X_i increases the value of Y?

Bias-Variance trade-off

To create a model, we split data into Training set+Test set

Bias-Variance trade-off

To create a model, we split data into Training set+Test set

Fix a data point (x_0, y_0) in the test set.
Treat $\hat{f}$ as **random** depending on the training set that we sampled from the population.

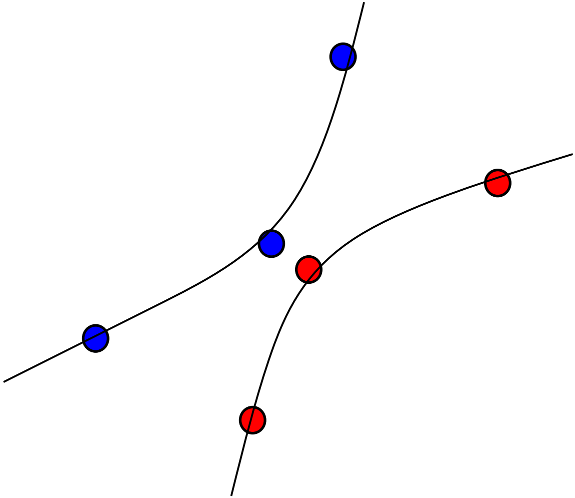
Bias-Variance trade-off

To create a model, we split data into Training set+Test set

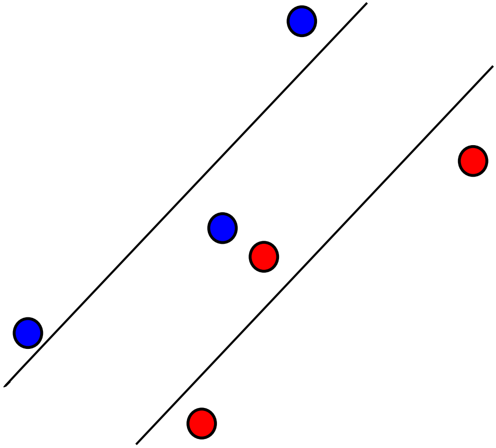
Fix a data point (x_0, y_0) in the test set.
Treat $\hat{f}$ as **random** depending on the training set that we sampled from the population. It can be shown that

$$\mathbf{E}\big(y_0 - \hat{f}(x_0)\big)^2 = \mathrm{Var}(\hat{f}(x_0)) + [\mathrm{Bias}(\hat{f}(x_0))]^2 + \mathrm{Var}(\epsilon).$$

Example



Example



Bias-Variance tradeoff