MONASH University

ETC3250

Business Analytics

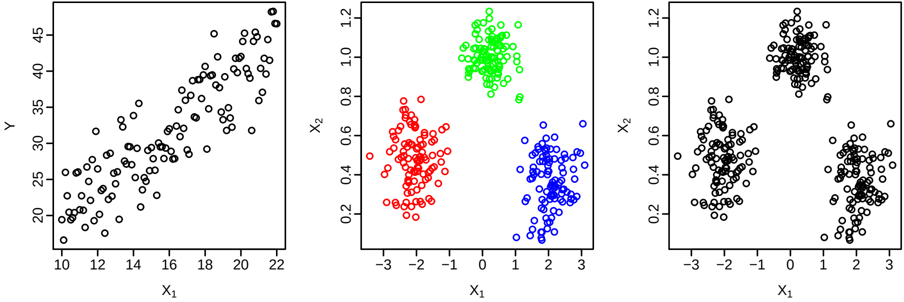
Week 2. Statistical learning

1 August 2016

Learning from data

- **Better understand** or **make predictions** about a certain phenomenon under study
- **Construct a model** of that phenomenon by finding relations between several variables
- If phenomenon is complex or depends on a large number of variables, an **analytical solution** might not be available
- However, we can **collect data** and learn a model that **approximates** the true underlying phenomenon

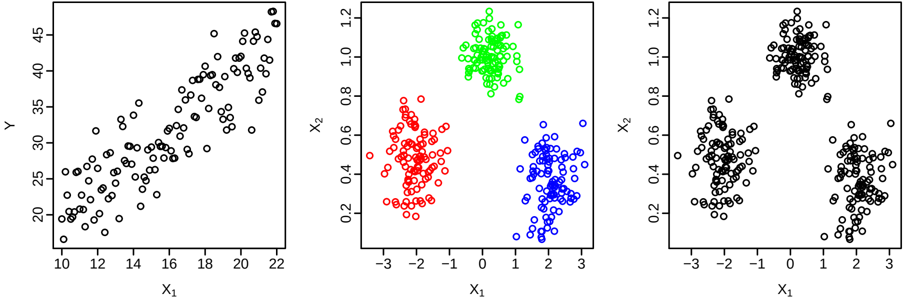
Learning from a dataset



$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N} \text{ with } x_i = (x_{i1}, \ldots, x_{ip})^T$$

Statistical learning provides a framework for constructing models from $\mathcal{D}$.

Learning from a dataset



$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N} \text{ with } x_i = (x_{i1}, \ldots, x_{ip})^T$$

Statistical learning provides a framework for constructing models from $\mathcal{D}$.

Different learning problems

- Supervised learning
  - Regression (or prediction)
  - Classification

  $\rightarrow y_i$ available for all x_i
- Unsupervised learning

  $\rightarrow y_i$ unavailable for all x_i
- Semi-supervised learning

  $\rightarrow y_i$ available only for few x_i
- Other types of learning: reinforcement learning, online learning, active learning, etc.

Identification of the best learning problem is important in practice

What is Statistical Learning?

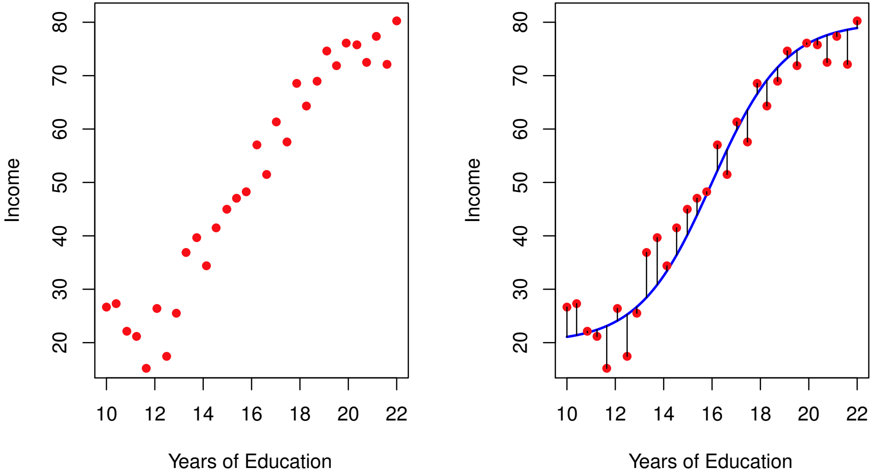
$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$$

$$Y = f(\overbrace{X_1, \ldots, X_p}^{X}) + \varepsilon$$

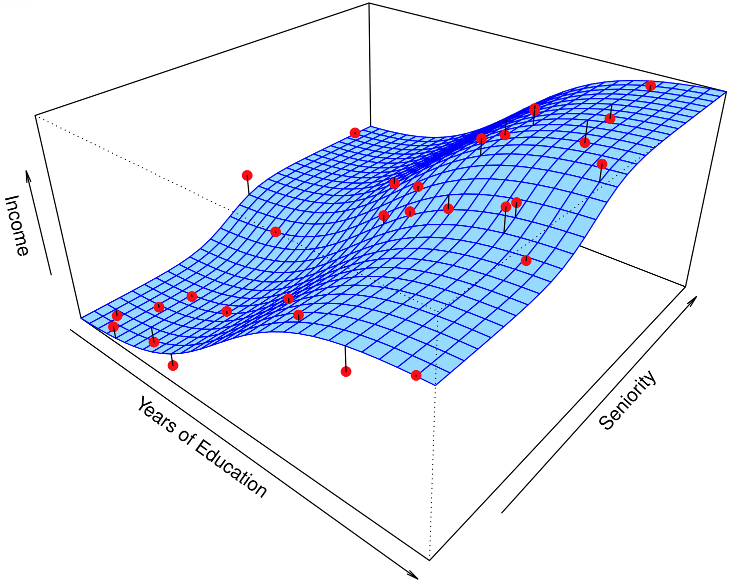
- Y: response (output)
- f: unknown function
- X: set of p predictors (inputs)
- ε: error term

Learn (or estimate) the function f using $\mathcal{D}$

What is Statistical Learning?



What is Statistical Learning?



Why estimate f?

- **Prediction**:
  - $\hat{Y} = \hat{f}(X)$

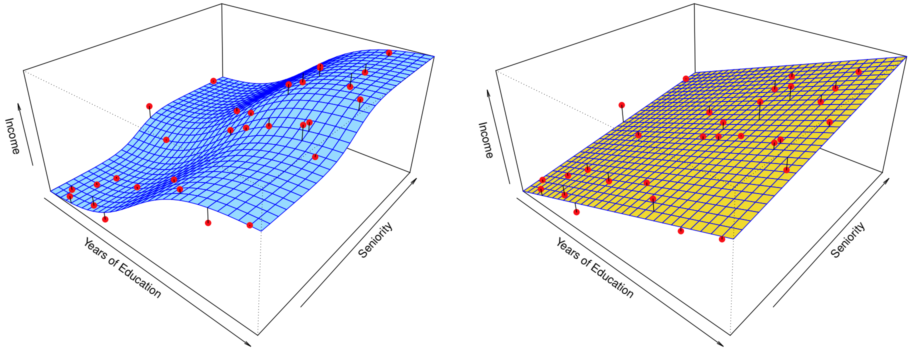
$$
\begin{aligned}
E[(Y - \hat{Y})^2] &= E[(f(X) + \varepsilon - \hat{Y})^2] \\
&= \underbrace{E[(f(X) - \hat{f}(X))^2]}_{\text{Reducible}} + \underbrace{\mathrm{Var}(\varepsilon)}_{\text{Irreducible}}
\end{aligned}
$$

- **Inference (or explanation)**:
  - Which predictors are associated with the response?
  - What is the relationship between the response and each predictor?

How do we estimate f?

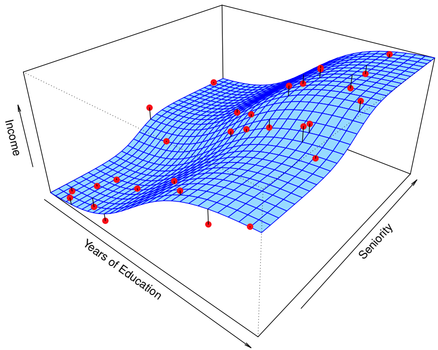
- **Parametric methods**
  - Assumption about the form of f, e.g. linear: $f(X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$ and $\hat{Y}(x) = \hat{f}(x)$
  - ☺ The problem of estimating f reduces to estimating a set of parameters
  - ☺ Usually a good starting point for many learning problems
  - ☹ Poor performance if linearity assumption is wrong
- **Non-parametric methods**
  - No *explicit* assumptions about the form of f, e.g. nearest neighbours: $\hat{Y}(x) = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i$
  - ☺ High flexibility: it can potentially fit a wider range of shapes for f
  - ☹ A large number of observations is required to estimate f with good accuracy

How do we estimate f?

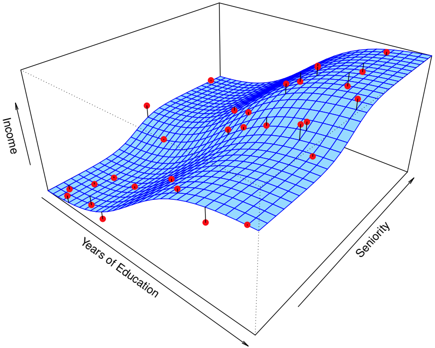


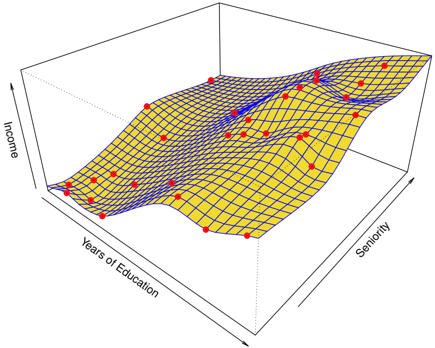
$$\hat{f}(\text{education}, \text{seniority}) = \hat{\beta}_0 + \hat{\beta}_1 \times \text{education} + \hat{\beta}_2 \times \text{seniority}$$

How do we estimate f?



How do we estimate f?





”Why would we ever choose to use a **more restrictive method** instead of a **very flexible approach**?”

Prediction Accuracy vs Model Interpretability