2/5/18

Lecture 3

Training Data

$\mathbb{D} = \{\langle \vec{x}_1, y_1 \rangle, \langle \vec{x}_2, y_2 \rangle, \ldots, \langle \vec{x}_n, y_n \rangle\}$

$\vec{x}_1$ is Bill's characteristics, $y_1 = 1$ means he paid for his loan

$\vec{x}_2$ is Jill's characteristics, $y_2 = 1$ means she paid for her loan

$\vec{x}_3$ is Tony's characteristics, $y_3 = 0$ means he did not pay his loan

or $\mathbb{D} = \langle X, \vec{y} \rangle$, $X \in \mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_p$

$\vec{y} \in \mathcal{Y}^n$

$$\underbrace{\begin{bmatrix} - \vec{x}_1 - \\ - \vec{x}_2 - \\ \vdots \\ - \vec{x}_n \end{bmatrix}}_{\text{size } n \times p} \quad \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\text{size } n}$$

Supervised Offline Learning

(Offline uses historical data)

You need 3 ingredients

① $\mathbb{D}$, the training data

② $\mathcal{H} := \{\text{all candidate functions for } f\}$

③ A, the algorithm which produces $g = A(\mathbb{D}, \mathcal{H})$

$y = t(\vec{z})$ → true causel inputs (unknown)

↑ phenomenon we wish to model

But $\vec{x}$ is observable

← approx to y

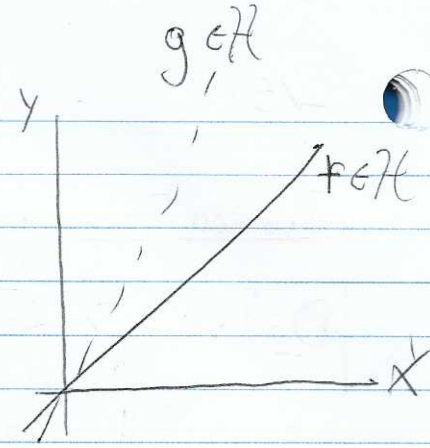
$y = f(\vec{x}) + \delta$, $\delta = t(\vec{z}) - f(\vec{x})$

($f(x)$ is both a true and a false function. It is false because it will never be reality. It is true because it uses the best available data we an use)

Goal: estimate f

g is best approx. to f in $\mathcal{H}$

If $f \in \mathcal{H}$ (unlikely) ...

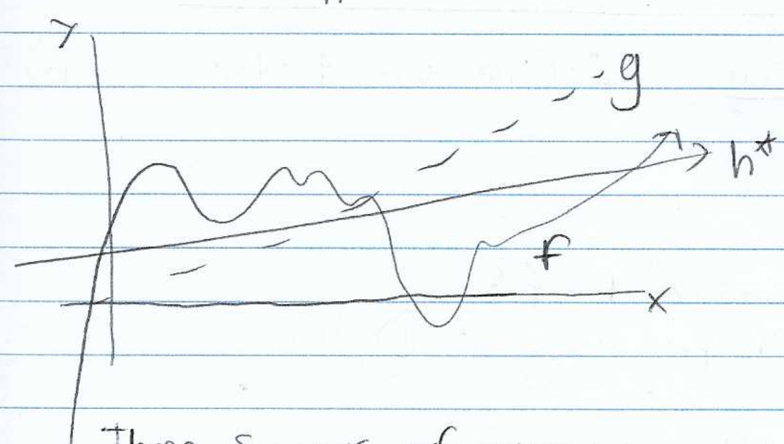


$$y = g(\vec{x}) + \underbrace{\left(f(\vec{x}) - g(\vec{x})\right)}_{\text{estimation error / parameter estimation error}} + \underbrace{\left(t(\vec{z}) - f(\vec{x})\right)}_{\delta:\ \text{error due to ignorance}}$$

If $f \notin \mathcal{H}$

$$y = g(\vec{x}) + \underbrace{\left(h^*(\vec{x}) - g(\vec{x})\right)}_{\text{estimation error}} + \underbrace{\left(f(\vec{x}) - h^*(\vec{x})\right)}_{\text{misspecification error}} + \underbrace{\left(t(\vec{z}) - f(\vec{x})\right)}_{\text{error due to ignorance}}$$

h^*: the best approx of $f \in \mathcal{H}$



Three Sources of error

To mimize estimation error you need a bigger n

misspecification error