01/31/ Assume $y = f(z_1, \ldots z_t)$

it is not the model, it is the reality.

Next best thing

obtain $X_1, \ldots, X_p$ which hopefully capture much of the information in the z's.

let $\vec{x}_i := [x_{i1}, x_{i2}, \ldots x_{ip}] \in \mathcal{X}$

$\vec{x}_i$: observation, subject, setting, record, objects

x_{ij}: variable, features, regressor, measurements

"input space"

X_1: Credit score $\in \mathbb{R}$

Continuous variable

X_2: criminality

many metrics.

$X_2 \in \{$ has past history, does not have $\}$

$\downarrow$ 1 $\quad$ $\downarrow$ 0

Binary Variable

Indicator Variable

Dummy Variable

$X_2 \in \{$ none, infraction, misdiminor, felony $\}$

factor variable or categorical Variable

w/ $L = 4$ levels.

now

Two strategies to use factor variables in mathematical models.

(a) ordinary encoding

$X_2 \in \{0, 1, 2, 3\}$

Ordinal factor variable

major downside: encoding is arbitrary $\Rightarrow$ you can assigne many ~~###~~ 1, 500 etc.

(b) nominal encoding

You create → $X_{2a} \in \{0,1\}$ infraction or not ?

$X_{2b} \in \{0,1\}$ misdemenor or not ?

$X_{2c} \in \{0,1\}$ felony or not ?

$X_{2a}, X_{2b}, X_{2c} = \emptyset$ "none"

downside: $p=3 \Rightarrow p=5$ more 1-1

there are things that we can not nominalize like favorite color "your choice"

Can we say: $y = f(x_1, \ldots x_p)$? No (not reality)

$y = f(x_1, \ldots x_p) + \delta$ → error to ignorance

δ: $t(\bar{z}) - f(x)$

error to thing we dont know even if $f(x..x)$ are identical

How to minimize δ:

increase # of relevant variables.

→ after finding x_s we need

↳ Find f. The approach we use is called "learning from data"

a "empirical approximation". (base on measurements, data)

The type of learning from data we will employ is "supervise learning"

Historical data overseas

We learn from measurement

buy → same amount

Supervised learning needs 3 ingredients.

① "training data", "historical data"

Examples of things come in and goes out similar

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

n: # of historical examples (sample size)

$\vec{x}_1$ is Bob's measurements $y_1 = 1$ (he paid)

$\vec{x}_2$ is Jill " $y_2 = 1$ (she paid)

$\vec{x}_3$ is Bill " $y_3 = 0$ (he did not paid)

$$X := \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_n \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \qquad \mathbb{D} = [X; \vec{y}]^{n \times p}$$

$n \times p$ ⇒ has p measurements not "1"

WE know nothing about f

e.g. $f: \mathbb{R}^p \to \mathbb{R}$ so we need to shrink the space down

assumption, because we not know f

② H = a set of candidate functions h that can approximate f.

③ A = an algorithm that takes H and $\mathbb{D}$ (→ historical data) and provides $g \in H$ as the best approximation of f.

$$g = A(\mathbb{D}, H)$$

Is $f \in \mathcal{H}$? Generally speaking .. no.

However $\exists$ (Exists) $h^* \in \mathcal{H}$ that is the best approximation of f.

$$y = h^*(x_1, \dots x_p) + \underbrace{f(\vec{x}) - h^*(\vec{x})}_{\text{missspecification error.}} + \underbrace{t(\vec{z}) - f(\vec{x})}_{\text{error we know we dont know to ignorance.}}$$

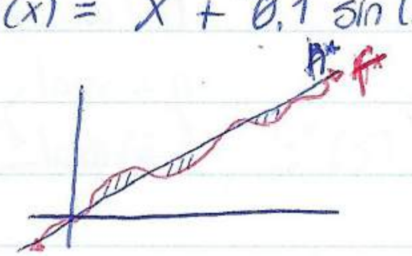
→ we can not minimize / get rid

this error is Captures in the graph

→ we can get rid of this error

→ you did not include enough functions in h^*

▸ Let $f(x) = x + \overbrace{0.1}^{\text{scale}} \sin(x)$



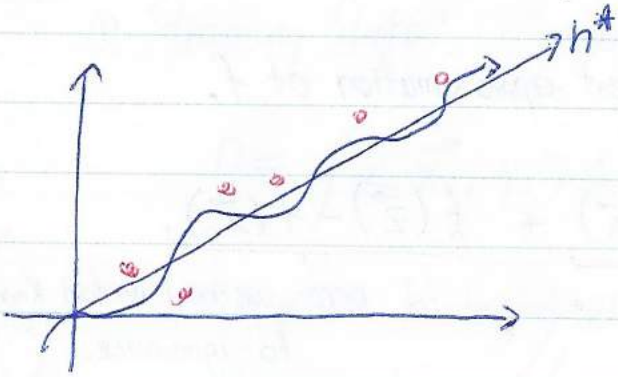
$\mathcal{H} = \{\text{all the linear functions of } x\}$
$= \{\beta_0 + \beta_1 x : \beta_0 \in \mathbb{R}, \beta_1 \in \mathbb{R}\}$
$h^*(x) = x$

now do dont have missespecification

will be good

$f(x) = x + 0.1 \sin(x)$
$\mathcal{H} = \{\text{all linear function of } x\}$
$= \{\beta_0 + \beta_1 x + \beta_2 \sin x : \beta_0, \beta_1, \beta_2 \in \mathbb{R}$

$\mathbb{D}$ = is 6 points in a piece paper



(h^*) our candidate

$g \rightarrow$ ideal
$f \rightarrow$ ideal → not know
} candidate

I want $g(x) = \underbrace{b_0 + b_1 x}_{\text{element of } H}$

$$y = g(\vec{x}) + \underbrace{h^*(\vec{x}) - g(\vec{x})}_{\text{stimation error}} + f(\vec{x}) - h^*(\vec{x})$$

$g \rightarrow$ is the only one we know.

$g(\vec{x})$: model

$\underbrace{\text{stimation error} + f(\vec{x}) - h^*(\vec{x})}_{\varepsilon \text{ (residual)}}$

How to predict?
↳ for a new object $\vec{x}_*$, how to predict y?

$$\vec{y} = g(\vec{x}_*)$$

$$y = h^*(\vec{x}) + \underbrace{f(\vec{x}) - h^*(\vec{x}) + t(\vec{z}) - f(\vec{x})}_{E \quad \text{Error (noise)}}$$

Rstudio

How to minimize missespecification Error?

Make H richer, A as well.

7 options.

ladder notatio

How to minimize stimation error?

Increase n (the sample size)

→ we can not know from where the error is comming from.

02/05

$$y = g(\vec{x}) + h^*(\vec{x}) - g(\vec{x}) + \varepsilon \text{ (residual)}$$

For a new observation x^*, $\hat{y} = g(x^*)$

g comes from In supervise learning → historical data

$$g = A(D, H)$$ → model space

A: algorithm

Loan Model

$y = \{0, 1\}$ → pay back loan (credit)

0: no pay back loan

→ Model is called binary classification model $\in \{0, 1\}$.

Null Model: You have no features and you were to create the best model g.

$H = y$ $g =$ ~~should~~ Model (y)