01/31/2019

Assumption: $y = f(z_1, \ldots, z_t)$ it is not the model; it is the reality.

Next Big Thing!

Obtain $x_1, \ldots, x_p$ which hopefully... captures much information in the z_s

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

- i: represents Bob; observation, setting, subject, object, input
- $[x_{i2}, x_{i2}, \ldots, x_{ip}]$: variance characteristic features regressor attributes covariance measurements.
- $\mathcal{X}$: input space, covariance space

x_1: credit score $\in \mathbb{R}$

continuous variables

x_2: criminality

many metrics

$x_2 \in \{$ has past history, does not have $\}$

indicator variable, binary variable, dummy...

1 0

$x_2 \in \{$ none, infraction, misdemeanor, felony $\}$

factor variable, categorical variable with $L=4$ levels

Two Strategies to use factor var's in math models.

(a) Ordinal encoding

misdemeanor felony

$x_2 \in \{0, 1, 2, 3\}$

ordinal factor variable.

major downside: encoding is arbitary

(b) Nominal encoding

$x_{2a} \in \{0, 1\}$ infraction or not?

$x_{2b} \in \{0, 1\}$ misdemeanor or not?

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

$x_{2a} = x_{2b} = x_{2c} = 0 \Rightarrow$ "none"

downside: too many inputs

$p=3 \Rightarrow p=5$ $\quad L-1=3$.

Mathematical Model

Can you say $y = f(x_1, \ldots, x_p)$?

No $\quad y = f(x_1, \ldots, x_p) + \underbrace{\delta}_{f(\vec{z}) - f(\vec{x})}$ → error due to ignorance

in two cases even this is identical but there still thing that you don't know.

How to minimize δ:
increase number of relevant variable. (income? governmen employee? marry or not?)

Find f. The approach we use is called "learning from data" (f) an "empirical approach" (based on data, measurement, observation); the type of learning from data we will employ is "supervised learning" (historical data oversee the learning).

Supervised Learning needs 3 ingredients

① "training data", "historical data" (one goes in something, similar comes out.)

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

n: # of historical exmples (sample size).

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

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

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

$$X := \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_n \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathcal{Y}^n$$

→ has dimmention $n \times p$

historical data $\mathbb{D} = [X, \vec{y}]$

eg. (Arbitrary, know nothing) $f : \mathbb{R}^p \to \mathbb{R}$

② $\mathcal{H}$:= a set of candidate functions h this can approximate f. (set of all linear function of x's)

Assumption!

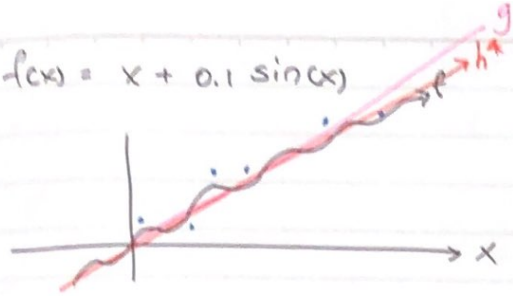
③ $\mathcal{A}$: an algorithm that takes $\mathcal{H}$ and $\mathbb{D}$ and provides $g \in \mathcal{H}$ as the best approximation of f. which is h^*.

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

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

However $\exists\, h^* \in \mathcal{H}$ this is the best approximation of.

$$y = h^*(x_1, \ldots, x_p) + \underbrace{f(\vec{x}) - h^*(\vec{x})}_{\text{misspecification error.}} + \underbrace{\epsilon(\vec{x}) - f(\vec{x})}_{\text{error due to ignorance}}$$



the more complex is the closer is to the line.

$\mathcal{H}$ = {all linear functions of x}

= $\{\beta_0 + \beta_1 x : \beta_0 \in \mathbb{R}, \beta_1 \in \mathbb{R}\}$

$h^*(x) = x$

want $g(x) = b_0 + b_1 x$

Example:

there are six data point

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

$g(\vec{x})$: Model.

$\underbrace{\quad\quad}_{e \text{ (residual)}}$

How to predict?

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

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

How to minimize misspecification error?

Make $\mathcal{H}$ richer, make A better?

How to minimize estimation error?

Increase n (sample size)