01/30/2020

$y = f(z_1, z_2, z_3)$ — f: true function; y: Phenomenon; z_1, z_2, z_3: true causal drivers

pay back mortgage (1 or 0)

$y \in y = [0, 1]$

z_1 : has the money (1)
z_2 : Unforces emergency (1)
z_3 : Is criminal (1)

Next Best thing: Find features that approximate the "information" in z_1, z_2, z_3.

Here are 3 that are directly related.
x_1 : Salary at time of application (continous) $\in \mathbb{R}$
x_2 : Miss previous payment (binary) $\in (0,1)$
x_3 : (criminal intent) Do they have a record? (binary) $\in (0,1)$

x_j's are called "features", "characteristics", "attributes", "regressors", "variable", "covariates", "independent variables"

Let p denote the #'s of features

let $\vec{x_i} := [x_{i1}, x_{i2}, \ldots, x_{ip}] \in x$ input space

$\vec{x_i}$ is called i^{th} ... "observation", "subject" "setting" "record" "input" "unit"

e.g $\mathbb{R}^p$

Can I measure X_3 'better' worst crime

$X_3 \in$ [none, infraction, misdemeanor, felony]

categorical or factor variable with $L = 4$ levels

Mathematical models required numerical values.

What do we do?

Two options.

① Note: this is R factor validated with a monotonic order code this variable

variable via $X_3 \in [0, 1, 2, 3]$

0 ↑ home, 3 ↑ felony

Downside :- The coding is arbitrary

* e.g why not

$X_3 \in [0, 1, 5, 100]$

② Create multiple features

$X_{3a} \in [0, 1]$ is infraction?

$X_{3b} \in [0, 1]$ is misdemeanor?

$X_{3c} \in [0, 1]$ is felony?

None is captured by $X_{3a} = X_{3b} = X_{3c} = 0$.

L-1 binary variables.

Consider $X_j \in$ [Red, Green, Blue]

unordered factors

$y = t(z_1, z_2, z_3) = f(x_1, \ldots, x_p) + f$

f - f is the error due to ignorance

f :- is the 'best' possible way of combining $x_1, \ldots x_p$ to minimize f

How to get f? ; Analytical solution? No,

Analytical solution for f.

$$f(x) = \int x\,dx = x^2/2$$

The approach we will use is "learning from data"
Use data to get an estimate of f.

This procedure is also termed "supervise learning".

There are three ingredients.

① Training Data (Data)

Notation; $\mathcal{D}$

$$\mathcal{D} = [(\vec{x_1}, y_1), (\vec{x_2}, y_2), \ldots (\vec{x_n}, y_n)]$$

n: Historical examples. (subject with response)
↓ sample size
i.e. - It happened already.

$\vec{x_1}$: Bob's features y_1: 1 (He paid it back)
$\vec{x_2}$: Bell's features y_2: 0 (He didn't pay back)

Standard notation:

$$X = \begin{bmatrix} \leftarrow \vec{x_1} \rightarrow \\ \leftarrow \vec{x_2} \rightarrow \\ \vdots \\ \leftarrow \vec{x_n} \rightarrow \end{bmatrix} \quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

col. vector length n.

nxp matrix

$$\mathcal{D} = [X, \vec{y}]$$

② H = a set of candidate function for f.

Recall $f: x \longrightarrow y$ e.g $F: \mathbb{R}^p \longrightarrow \mathbb{R}$
You need to simplify the set of possible function.

③ A: an algorithm; such that $g = A(\mathbb{D}, H)$ away to select/learn a model. $g \in \mathcal{F}$ using ①