Recall: $y \in [0,1] = Y$

The system (not a model but some collect:

$$Y = t(z_1, z_2, z_3)$$

to z_1 = has sufficient funds

z_2 = Unforseen emergency

z_3 = Criminal mind

$$Y = z_1(1-z_2)(1-z_3)$$

Big problem: $\{z_1, z_2, z_3\}$ are unobserable.

What to do?

Reason: We are trying to get something that related to $\{z_1, z_2, z_3\}$.

Say: Impossible: get $\{z_1, z_2, z_3\}$, t

Next test: Try to define and collect info "related to the true crucial impos.

- let X_1: Salary

How to measure it?

Maybe: X_1: Salary: How to measure Avg over 5yr related to z_1

We are trying to design something to capture z_2.

- X_2: previous loan repayment $\in [0,1]$
- X_3: previous crime type {no crime, infract:, felony}

Process assessment: use much you got, use much you have ~~exple, gender and age~~, use what is cheaply obtainable.

~~Exple~~: use age,

Let say we have X_1, X_2 and X_3.

The idea is: $\{z_1, z_2, z_3\}$

~~$\{X_1, X_2, X_3\}$~~ ?? contain some of info

$\{X_1, X_2, X_3\}$

Bob's info: ← is an "observation", "Record", "object", "input", indepdt Varia

$\vec{X} = [X_1, X_2, X_3]$

$\dim(\vec{X}) = p$ or "d"

X_1, X_2, X_3 are features of Bob's
are characteristics, attributes
in economic we call it Regressors,
covariance, predictors.

$\vec{X} = [X_1, X_2, X_3] \in \mathcal{X}$ " covariance space.

$X_1 \in \mathbb{R}$

$X_2 \in [0,1]$: binary or dinary variable.

X_3 is categorical variables with 4 "levels"
↙ imagine possible Valu

Two Ideas

(A) First to do: Code is numerically

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

it should only be done if predicator is "ordinal"

(B) Next to do:

Take X_3 & turn it to X_{3a} X_{3b} X_{3c} X_{3d}

- X_{3a} → binary no crime
- X_{3b} → binary
- X_{3c} → binary salary
- X_{3d} → binary felony

X_3 = misdemeanor $\vec{x} = [0 \ldots 0, 0]$

$p = 3 \longrightarrow 6$

So, it is impossible to get $\{z_1, z_2, z_3\}$ but we do have $[x_1, x_2, x_3]$

GOAL: do the best we can to explain y by creating model f, the approximation, the best real relationship we can get.

Does $y = f(x_1, x_2, x_3)$? : No

Note: I assumed, $y \approx f(x_1, x_2, x_3)$

$$y = f(x_1, x_2, x_3) + \delta$$

where $\delta = t(\vec{z}) - f(\vec{x})$

↖ comes from ignorance.

How do we get f?

First note there is no analytical solution.

exple: $h(x) = x^2$ Find $\min\{h\}$

Instead use an "empirical solution",

exple use data $\Rightarrow$ "Learning from data"

"Supervised Learning"

uses historical exples of record & their responses.

In this case, it requires 3 ingredients:

① $\mathbb{D} := \{(\langle \vec{x}_1, y_1 \rangle, \langle \vec{x}_2, y_2 \rangle, \langle \vec{x}_3, y_3 \rangle\}$

$\vec{x}_1$ is Billis characteristics and y_1 is ~~solut~~ whenever not paid but loan.

$\vec{x}_2$ is Jill ----------

Let $X = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_n \end{bmatrix} \in \mathcal{X}^n$.

$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathcal{Y}^n$

$\dim(X) = n \cdot p$

$\dim(\vec{y}) = n$