1/31/18 Lecture 2

$y \in \{0,1\} = \mathcal{Y}$ output space

true system (not a model... but it's sometimes called the "true model")
It's not a model because it IS reality

z_1 = has sufficient funds
$y = t(z_1, z_2, z_3)$ z_2 = unforseen emergency
z_3 = criminal intent

Big Problem: $\{z_1, z_2, z_3\}$ are unobservable

Impossible: get $\{z_1, z_2, z_3\}$, t
Next Best: try to define and collect information "related" to the causal inputs.

- x_1: salary. How ~~to~~ to measure: avg. over 5 yr.
- x_2: previous loan repayment $\in \{0,1\}$ ever defaulted
- x_3: previous crime type {no crime, infraction, misdimeanor, felony}
- (Very difficult to obtain these xs) (x_3 not #'s)

Business-as-usual: Use what you have or use what is cheaply obtainable

($\{z_1, z_2, z_3\} \approx$ some of the same info $\approx \{x_1, x_2, x_3\}$)

__Bob's information:__

$\vec{x} = [x_1, x_2, x_3]$ → "covariates", "predictors", "features", "variables", "characteristics", "attributes", "regressors"
→ "observation", "record", "object", "input", "independent variables"

$\vec{x} = [x_1, x_2, x_3] \in \mathcal{X}$ "covariate space"

$x_1 \in \mathbb{R}$ "continuous variable"
$x_2 \in \{0,1\}$ "binary or dummy variable"
x_3 is a "categorical variable" with 4 "levels" (unique possible values)

(A) Code it numerically $x_3 \in \{0,1,2,3\}$ This should only be done if the categorical predictor is "ordinal"

(B) Dummification

$X_3 \to X_{3a}, X_{3b}, X_{3c}, X_{3d}$

X_{3a}: binary, no crime
X_{3b}: binary, infraction
X_{3c}: binary, misdemeanor
X_{3d}: binary, felony

x_3 = misdemeanor $\vec{x} = [\cdot \cdot\ 0\ 0\ 1\ 0]$

$p = 3 \to 6$

Goal: do the best we can in explaining y by creating model f, the approx., i.e. the best functional relationship we can get.

Does $y_{\text{ever}} = f(x_1, x_2, x_3)$? No...

Instead, $y \approx f(x_1, x_2, x_3)$

$y = f(x_1, x_2, x_3) + \delta$ $\quad \delta = t(\vec{z}) - f(\vec{x})$ comes from ignorance

How do we get f?

First note there is no analytical solution e.g. $h(x) = x^2$ find min h.
Instead use an "empirical solution" ie. use data.
$\Rightarrow$ "Learning from data"
"Supervised Learning" uses historical examples of records and their respo

It requires 3 ingredients:

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

$\vec{x}_1$ is Bill's characteristics and y_1 is whether or not he paid back loa

$\vec{x}_2$ is Jill's " " y_2 " "

Let $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathcal{X}^n$ $\quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathcal{Y}^n$ $\quad \dim(x) = n \times p$, $\dim(y) = n$