1/31/2018

2

- ~~Last Time~~: Continued...
  - $y \in \{0,1\} = Y$ output space.
  = true system (not a model, but it's sometimes called the "true model").
  $y = t(z_1, z_2, z_3)$, where z_1 = has sufficient funds, z_2 = unforseen emergency, z_3 = criminal intentions.
  - Say $y = z_1(1-z_2)(1-z_3)$
  - Big Problem: z_1, z_2, z_3 are unobservable.
    Also, we can't tell the future.
  - So, impossible to get $\{z_1, z_2, z_3\}$, t ⇒
    Next Best: Try to define and collect information "related" to the true causal inputs:
    x_1 = Salary: How to measure? Avg. over yr.
    x_2 = Previous loans repayments $\in \{0,1\}$ never defaulted
    x_3 = Previous crime type {no crime, infraction, misdemeanor, felony}. at least one?
  - Business-as-usual: Use what you have or use what is cheaply dattable.
- Example: Bob's information:
  $\vec{x} = [x_1, x_2, x_3] \in \mathcal{X}$, $dim(\vec{x}) = 3$ aka p or d.

  $\vec{x}$ ↑ observation, record, object, input "independent variables"

  ← features, variables (dependent), characteristics, attributes, regressors, covariates, predictors

  → $\mathcal{X}$ is called "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.
- How do we deal with this?
  Some ideas:
  a) Code it numerically, e.g.
     $x_3 \in \{0,1,2,3\}$
     → but this should only be done if the categorical predictor is "ordinal."
  b) Dummification:
     $x_3 \to x_{3a}, x_{3b}, x_{3c}, x_{3d}$
     binary binary
     e.g. x_3 = misdemeanor $\vec{x} = [\cdots 0\ 0\ 1\ 0]$
     → Problem: Now we have more (dimensions) covariates, i.e. p is now 6.
- So, 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 in explaining y by creating a model f, the approx; i.e. the best functional relationship we can get.
  - Does $y = f(x_1, x_2, x_3)$ ever? No...
    Instead, $y \approx f(x_1, x_2, x_3)$ ⇒
    $y = f(x_1, x_2, x_3) + \delta$, where
    $\delta = t(\vec{z}) + t(\vec{x})$.
    comes from
  - How do we get f?
    → First note there is no analytical solutions.
    e.g. $h(x) = x^2$ Find $\min(h)$.
    → Use an "empirical solution" i.e. use data,
    ⇒ "Learning from Data!"

- "Supervised learning:" Uses historical examples of records & their responses. It requires 3 ingredients:

(I) $\mathbb{D} = \{\langle \vec{x}_1, y_1 \rangle, \langle \vec{x}_2, y_2 \rangle, \ldots, \langle \vec{x}_n, y_n \rangle\}$, where $\vec{x}_1$ is Bill's characteristics and y_1 is whether or not he paid back the loan.

$\vec{x}_2$ ... Jill's ..., y_2 " " ..., etc.

(II) $\mathbb{D} = \langle X, \vec{y} \rangle$, let $X = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_n \end{bmatrix} \in \mathcal{X}^n$ &

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

$\Rightarrow \dim(X) = n \cdot p$ & $\dim(\vec{y}) = n$.

- goo.gl/ig8sV6

Github.com/gd6