Math 390.4
Lecture 2

$y = t(z_1, z_2, z_3)$
↑ Funct. ↑ true causal values/drivers
phenomenon
pay back mortgage (0 or 1)

$y \in Y = \{0, 1\}$

$z_1$: has the money (1)
$z_2$: unforseen emergency (1)
$z_3$: criminal intent (1)

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

Next best thing
Find features that approximate the information in $z_1, z_2, z_3$

$x_1$ = salary at time of application (continuous) $\in \mathbb{R}$
$x_2$ = missed previous payment (binary) $\in \{0, 1\}$
$x_3$ = Do they have a criminal record? (binary) $\{0, 1\}$

The $x_j$'s are called: features, characteristics, attributes, variable, independent variables, regressors, covariates

Let p denote the number of features
let $\vec{x}_i = [x_{i1}, x_{i2}, \dots x_{ip}] \in \mathcal{X}$ input space
eg: $\mathbb{R}^p$
$\vec{x}_i$ is called the $i^{th}$ observation, setting, record, unit, subject, input

improving the $x_j$'s

$x_3 \in \{$none, infraction, misdemeanor, felony$\}$
* categorical or factor variable with L=4 levels
Mathematical models require numerical values
Two options:
1) code them with variable num (monotonic order)
$x_3 = \{0, 1, 2, 3\}$
Downside: Arbitrary coding
$x_3 \in \{0, 1, 5, 100\}$
2) create multiple features
$x_{3a} \in \{0, 1\}$ ; is infraction?
$x_{3b}$ ; is misdemeanor?
$x_{3c}$ ; is felony?
no none bc it is captured in
$x_{3a} = x_{3b} = x_{3c} = 0$

consider:
$x_{ij} \in \{$Red, green, blue$\}$
↑ unordered factor
→ option 2

note: $y = t(z_1, z_2, z_3) \neq f(x_1, x_2 \dots x_p)$
$= f(x_1, x_2 \dots x_p) + \delta$
$\delta := t - f$ is the error due to ignorance (delta)
$f$: the best possible way of combining $x_1, \dots, x_p$ to minimize $\delta$

How to get f?
X Analytical solution
✓ Learning from data
↳ use data to come up with an estimate for $f$
↳ "supervised learning"

3 ingredients for "supervised learning"
1) Training data (data)
$\mathbb{D} = \{<\vec{x}_1, y_1>, <\vec{x}_2, y_2>, \dots, <\vec{x}_n, y_n>\}$
n historical examples (subject with response)
n: sample size ↑ happened already

$\vec{x}_1$: Bob's features
$y_1 = 1$ (he paid it back)
$\vec{x}_2$ = Bell's features
$y = 0$ (she did not pay it back)

- standard Notation

$X = \begin{bmatrix} \leftarrow \vec{x}_1 \rightarrow \\ \leftarrow \vec{x}_2 \rightarrow \\ \vdots \\ \leftarrow \vec{x}_n \rightarrow \end{bmatrix}$ : $n \times p$ matrix

$\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ col. vector length n

$\mathbb{D} = \{X, \vec{y}\}$

a) $\mathcal{H}$ : a set of candidate function for $f$
recall $f: X \to y$ "eg: $f: \mathbb{R}^p \to \mathbb{R}$"
Y ↑ → X
You need to simplify the set of possible functions

3) $A$ is in an algorithm
s.t. $g = A(\mathbb{D}, \mathcal{H})$
a way to select a model
$g \in \mathcal{H}$ using $\mathbb{D}$