MATH 390.4 Lecture 2

Pizon Shetu

January 30 2020

Note: Half the class was spent on Lab work e.g. RStudio overview which is on class github.

1 Features

$$Y = t(Z1, Z2, Z3)$$

Y was the phenomenon of pay back mortgage $\mathrm{Y} \in 1, 0$

$y \in Y = 0, 1$

$tisthefunction, Z'sarethetruecausaldrivers$

Z1: has the money (1) $\mathrm{Z1} \in 1, 0$

$Z2 : unforeseenemergency(1)Z2 \in 1, 0$

$Z3 : criminalintent(1)Z3 \in 1, 0$

$t = Z1(1 - Z2)(1 - Z3)$

Next best thing: Find features that approximates the "information" in Z1,Z2,Z3. Here are three that are directly related

X1: Salary at the time of application

$$X1 \in R+$$

X2: Payment history (Binary)

$$X2 \in R+$$

X3: Criminal History/ do they have a record? (Binary) X3 $\in 0, 1$

Xj's are called the following: features, characteristics, attributes, variables, independent variables, regressor, co-variance.

Let P := the number of features

Let Xi := [Xi1,Xi2,.....,Xip] element of X input space e.g. R^P

Xi is called the ith: observation, setting, record, input, unit, subject

Can I measure X3 better?

X3 $\in none, infraction, misdemeanor, felony$

This is a categorical or factor variable with L = 4 levels Mathematical models require numerical values

Two options: Note: this is a factor variable with a monotonic order. Code this variable in X3 = 0,1,2,3, but this has a downside, which is the coding is arbitrary e.g why not.... X = 0,1,5,100

Create multiple features: X3a $\in 0, 1 is infraction? X3b \in 0, 1 is misdemeanor? X3c \in 0, 1 is felony? None is not required since it is captured by X3a = X3b = X3c = 0 L-1 binary variable$

Consider Xj element of Red, Green, Blue (un-ordered factor) =¿ This must use option two

$$Y = t(Z1, Z2, Z3)! = f(X1.....Xp)$$

BUT

$$Y = t(Z1, Z2, Z3) = f(X1.....Xp) + \delta$$

$\delta := t - f is the error due to ignorance f := is the \backslash best" possible way of combining X1. . . . .Xp to minimize \delta.$

How to get f? Is there an analytical solution? No The approach we use is "learning from data". We use data to get an estimate of an f. This procedure is also called "supervised learning"

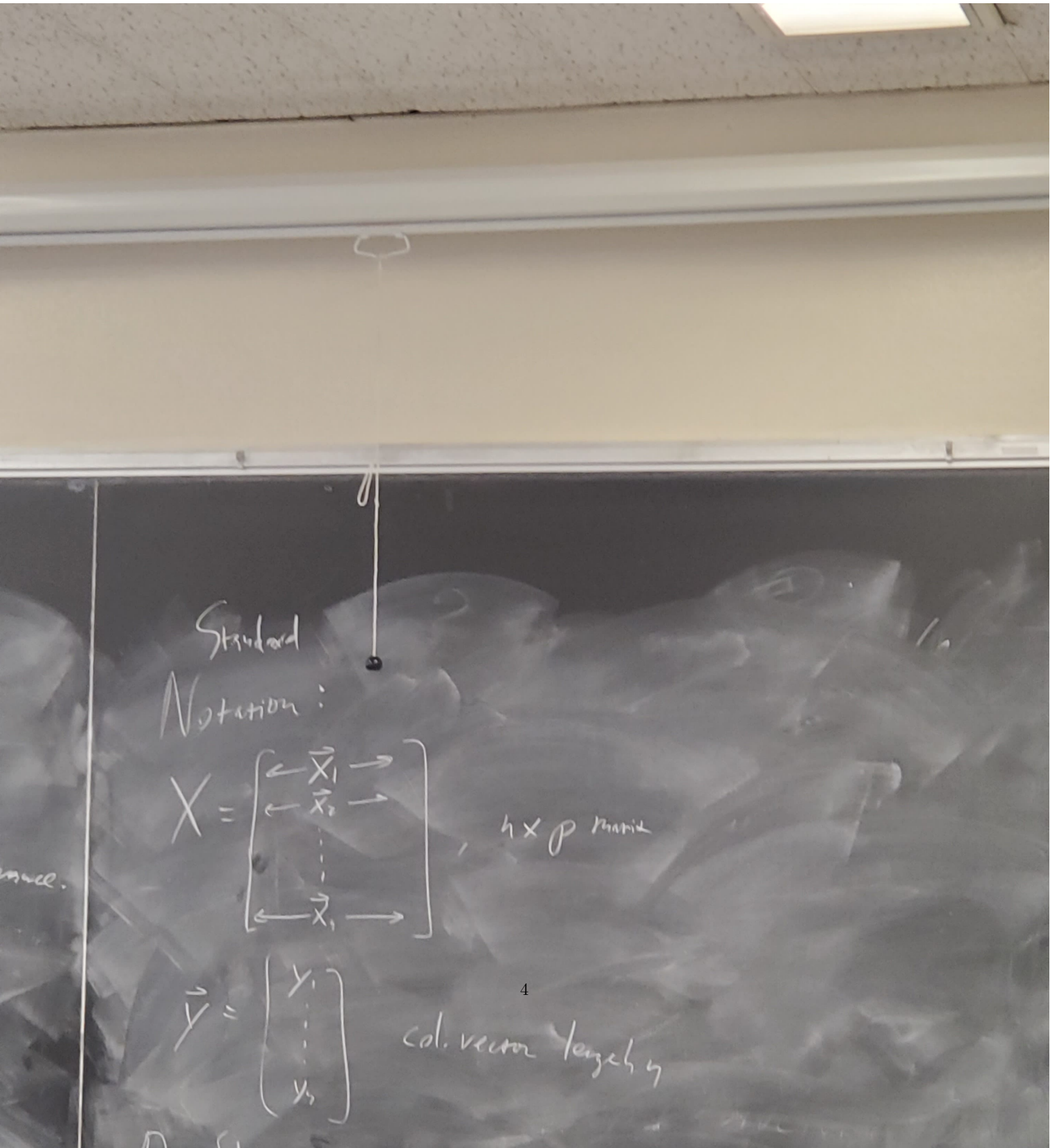
There are 3 ingredients: //

1. Training Data (Data) :=

$$D = < x1, y1 >, < x2, y2 > ..... < xn, yn >$$

This is n *(n is sample size) historical examples (subjects with response) i.e it happened already // x1: Bob's features // y1: 1 (he paid it back) // x2: Bills features // y2: 0 (He did not pay back his loan) //

Fig 2:


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

2. H = a set of candidate function for f, recall that f: R^P -¿ y e.g. f: R^P -¿ R You need to simplify the set of all possible functions

3. A is an algorithm: s.t. g = A(D,H), a way to select a model g which is element of H using D.