

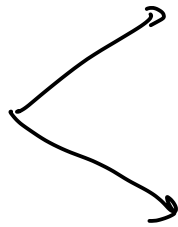
CS 307

FALL 2023

DALPIAZ

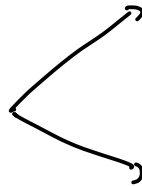
WEEK 02

ML



SUPERVISED

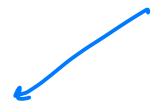
UNSUPERVISED



REGRESSION

CLASSIFICATION

WE ARE
HERE!



REGRESSION

DISTRIBUTION OF Y FOR SOME X

$$Y | X = x$$

← WANT TO LEARN

$$\mathbb{E}[Y | X = x]$$

MEAN OF Y FOR SOME X

DATA

X

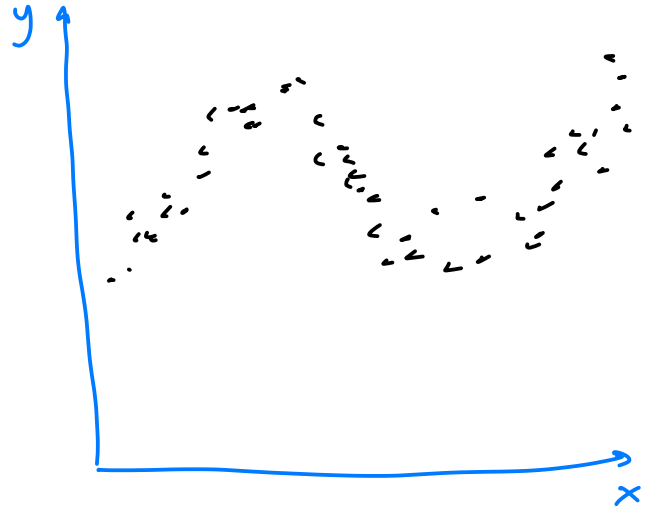
y

Numeric

x_1	x_2	x_3
...
...
...
...

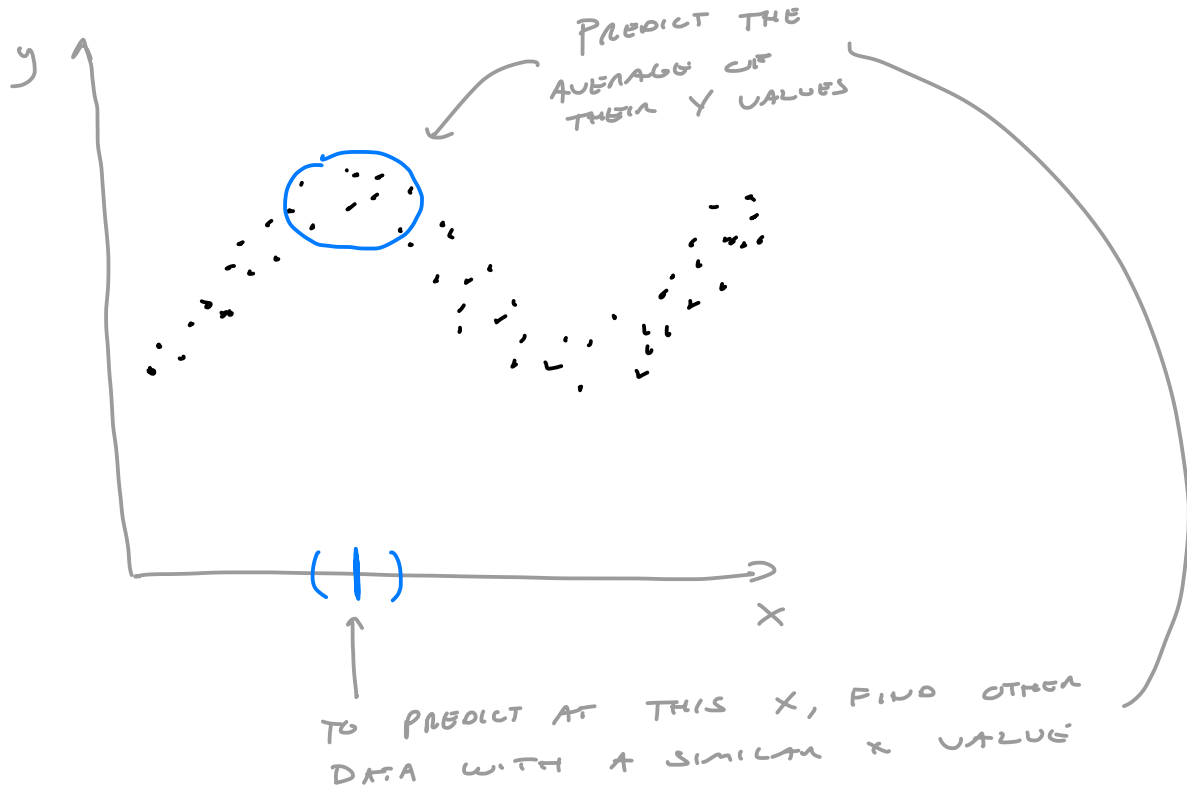
y
...
...
...
...

"TABULAR VIEW"



"GRAPHICAL VIEW"

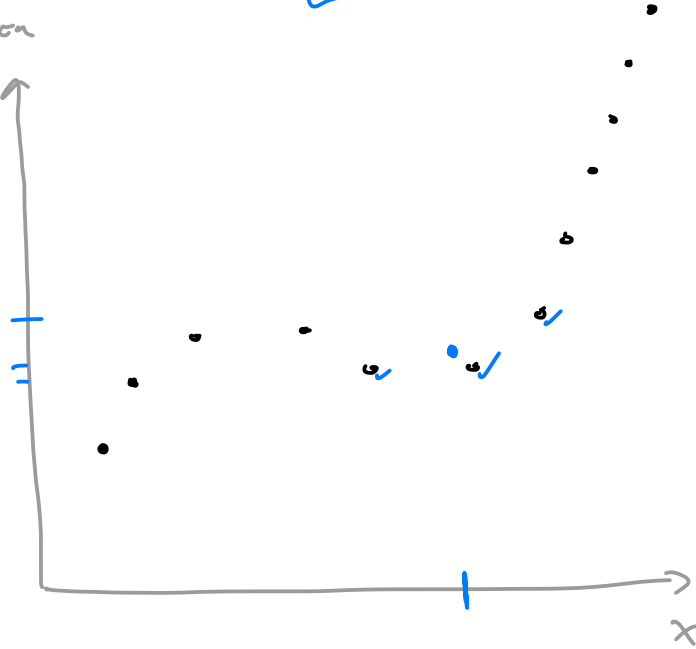
INTUITION



k-NEAREST NEIGHBORS

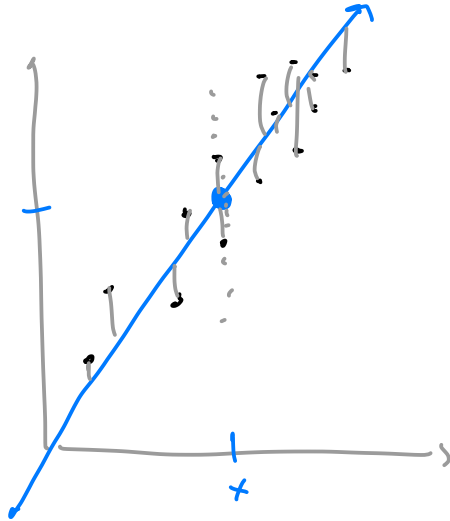
$k=3$

0. Fix k ← TUNING PARAMETER
1. New x
2. FIND THE k -NEAREST NEIGHBORS
↳ WRT x
3. AVE y 's OF NEIGHBORS



NON PARAMETRIC

LINEAR REGRESSION



$$Y = \underbrace{\beta_0 + \beta_1 x}_{\text{SIGNAL}} + \underbrace{\varepsilon}_{\text{noise}}$$

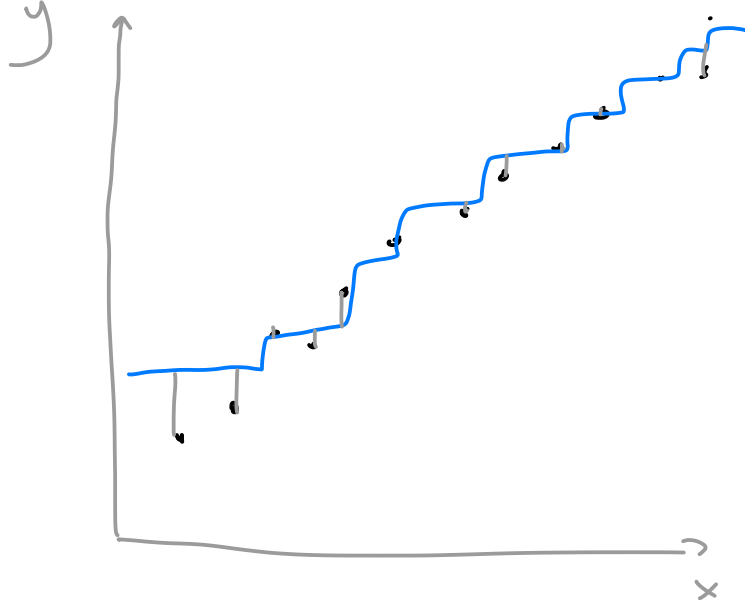
MODEL PARAMETERS

$$\varepsilon \sim N(0, \sigma^2)$$

LINEAR
IND
NORMAL
ERROR

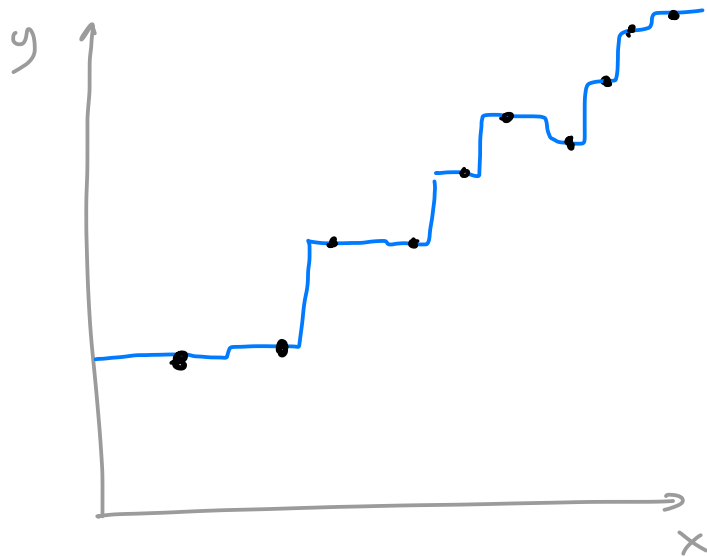
PARAMETRIC

$K=3$



ROOT MEAN SQUARE ERROR (RMSE) = $\sqrt{\sum_{i=1}^n (y_i - \hat{f}(x_i))^2}$

$$K=1$$

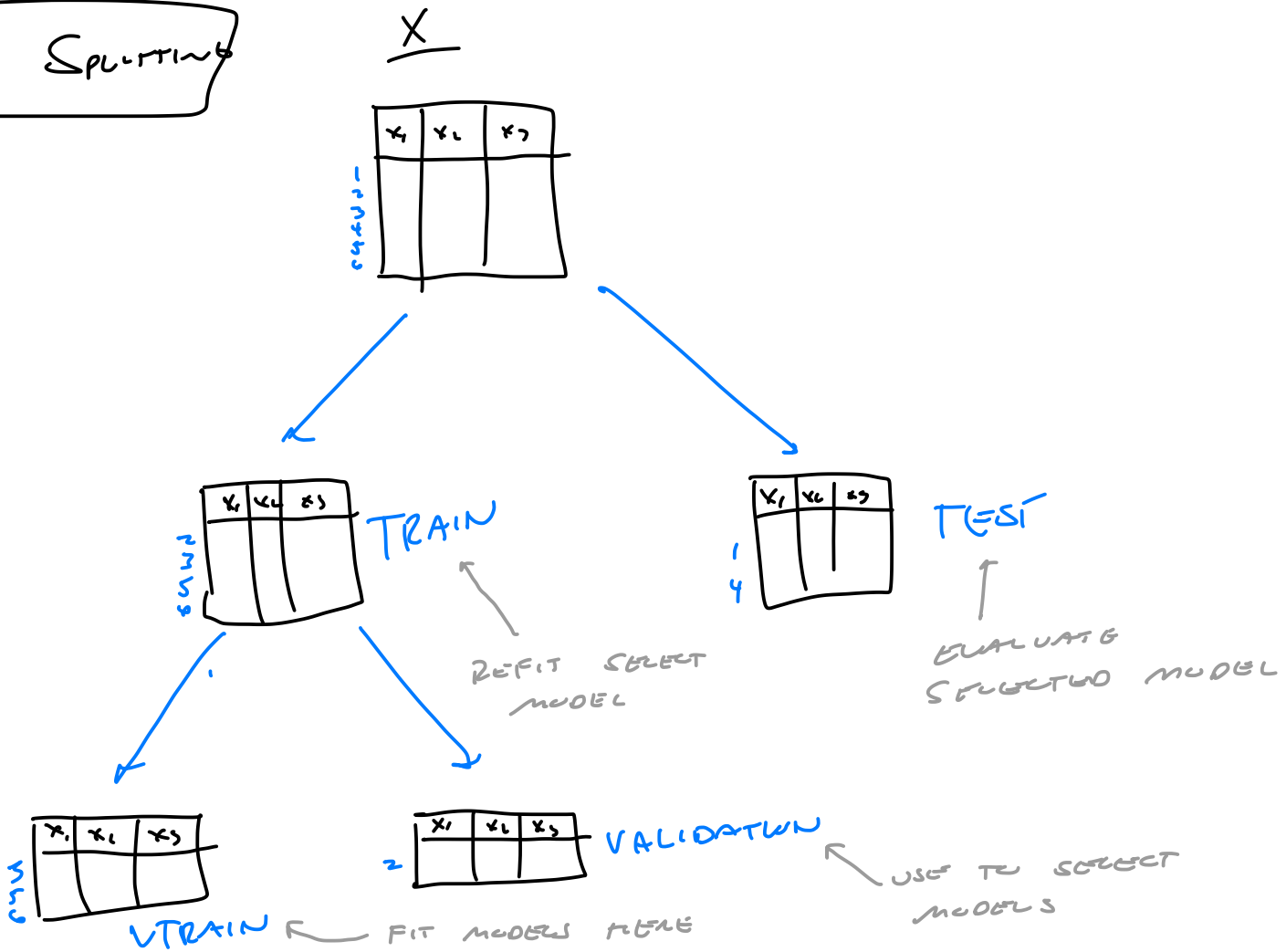


No ERRORS!

GENERALIZATION

A MODEL GENERALIZES WELL IF
IT IS ABLE TO MAKE GOOD PREDICTIONS
FOR "NEW DATA"

DATA SPLITTING



$$Y = \underbrace{f(x)}_{\text{SIGNAL}} + \epsilon$$

↑
LEARN THIS

↑
NOISE

↑
GET THIS
IF YOU DO, YOU
HAVE "LOSS"