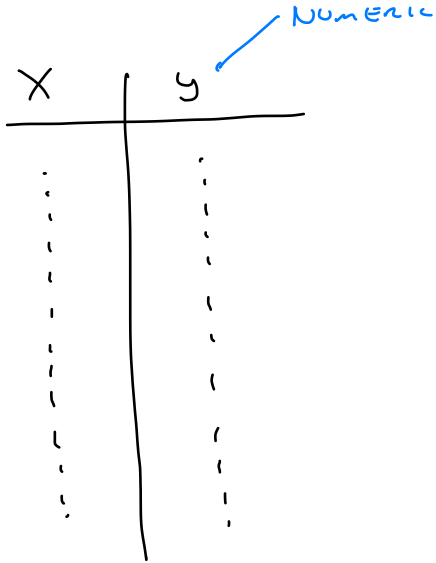
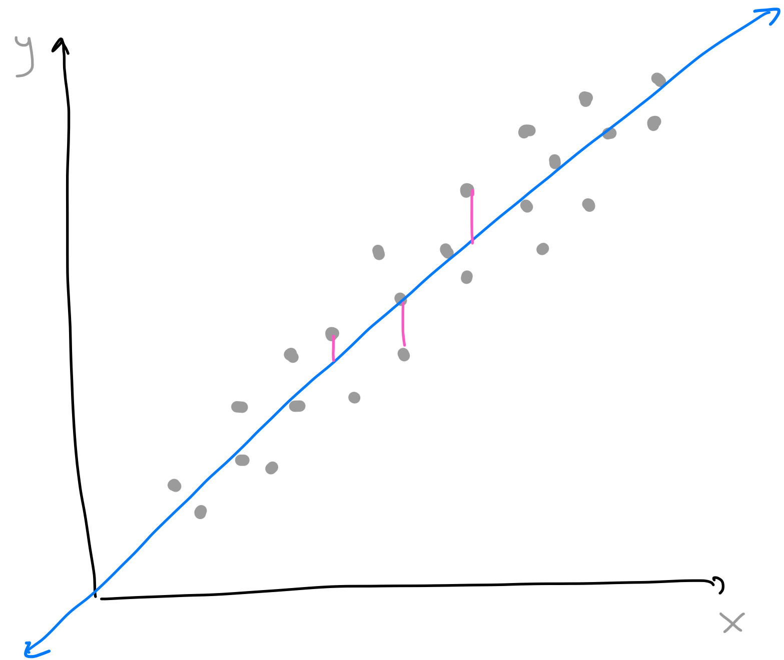
Linear Regression

CS 307


X
y
Numeric


y
X

GOAL OF REGRESSION, FOR ML:

MAKE GOOD PREDICTIONS ABOUT
A NUMERIC TARGET VARIABLE
FOR NEW INPUT FEATURE DATA

STATISTICAL GOAL OF REGRESSION ?

CONDITIONAL

TO LEARN THE **DISTRIBUTION**

OF Y GIVEN X

OFTEN SETTLE FOR LEARNING:

$$\mu(x) = \mathbb{E}[Y \mid X = x]$$

CONDITIONAL MEAN OF Y GIVEN x

TARGET

FEATURES

$$Y = f(X) + \varepsilon$$

SIGNAL

NOISE

WANT TO LEARN

NONPARAMETRIC METHODS

- USE "CLOSENESS" OF DATA
- MAKE FEW IF ANY ASSUMPTIONS

PARAMETRIC METHOD

- MAKE STRONG ASSUMPTIONS, ESPECIALLY ABOUT f

LINEAR MODEL ASSUMPTIONS

LINEARITY → DOES NOT IMPLY YOU CAN ONLY FIT A "LINE"

INDEPENDENCE

~~NORMAL~~

~~EQUAL~~

NEED FOR INFERENCE OR PREDICTION INTERVALS

$$Y = \underline{f(X)} + \varepsilon$$

WHAT ASSUMPTIONS ARE MADE ABOUT f?

$$Y = \underline{\beta_0 + \beta_1 x} + \varepsilon,$$

Assume $f(x) = \underline{\beta_0 + \beta_1 x}$

LINEAR FUNCTION OF x

UNKNOWN SLOPE (β_1)

AND INTERCEPT (β_0)

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

"ERRORS" OFTEN ASSUMED NORMAL WITH CONSTANT VARIANCE

MORE COMPACTLY

$$Y \mid X = x \sim N(\underline{\beta_0 + \beta_1 x}, \sigma^2)$$

$$\beta_0 + \beta_1 x = \mathbb{E}[Y \mid X = x] = \mu(x)$$

FITTING LINEAR REGRESSION

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

← MODEL TO FIT

DATA TO FIT WITH

x	y
⋮	⋮

$(x_i, y_i) \quad i = 1, \ldots, n$

$$\min_{\mu} \sum_{i=1}^{n} \left(y_i - \mu(x_i)\right)^2$$

$$\mu(x) = \beta_0 + \beta_1 x$$

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

"LEAST SQUARES"

↓

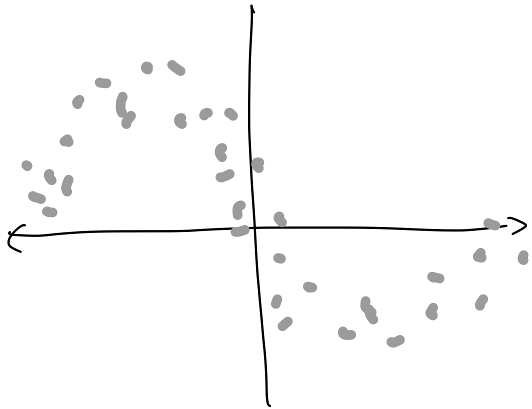
OPTIMIZATION HAPPENS

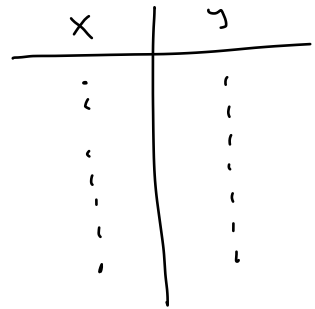
↓

$$\underline{\hat{f}(x)} = \hat{\beta}_0 + \hat{\beta}_1 x$$

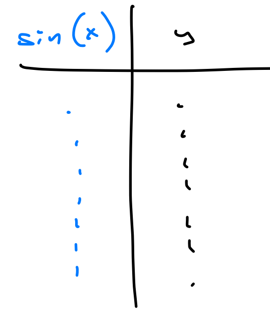
LEARNED f

LEARNED PARAMETERS





FEATURE ENGINEERING



ORIGINAL DATA

LINEAR MODEL UNABLE TO LEARN TRUE RELATIONSHIP

TRANSFORMED DATA

LINEAR MODEL CAN LEARN TRUE RELATIONSHIP

Fit $Y = \beta_0 + \beta_1 x + \varepsilon$ . . .

REPLACE WITH $\sin(x)$

- TRANSFORMATIONS ARE NOT ALWAYS OBVIOUS
- NONPARAMETRIC METHODS OFTEN WOULD NOT NEED THESE TRANSFORMATIONS

Multiple Linear Regression

x_1	x_2	x_3	y
$\vdots$	$\vdots$	$\vdots$	$\vdots$

$$Y = \boxed{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3} + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

$$\mu(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

"LINEAR COMBINATION"

NESTED MODELS

$$\mu(x) = \beta_0$$

$$\mu(x) = \beta_0 + \beta_1 x_1, \quad \mu(x) = \beta_0 + \beta_2 x_2, \quad \mu(x) = \beta_0 + \beta_3 x_3$$

$$\vdots$$

$$\mu(x) = \beta_0 + \underbrace{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}_{\text{"FIRST ORDER"}} + \underbrace{\underbrace{\beta_4 x_1^2 + \beta_5 x_2^2 + \beta_6 x_3^2}_{\text{QUADRATIC}} + \underbrace{\beta_7 x_1 x_2 + \beta_8 x_2 x_3 + \beta_9 x_1 x_3}_{\text{INTERACTION}}}_{\text{"SECOND ORDER"}}$$

PARAMETRIC

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

MODEL PARAMETERS

LEARNED FROM DATA

NON PARAMETRIC

KNN, $K = 3$

TUNING PARAMETER

DEFINE HOW TO LEARN

CHOSEN VIA A VALIDATION PROCESS

HOW TO INCREASE/DECREASE FLEXIBILITY?

- USE MORE/LESS FEATURES
- ADD TRANSFORMATIONS

LINEAR REGRESSION IN PYTHON WITH

sklearn.linear_model.LinearRegression