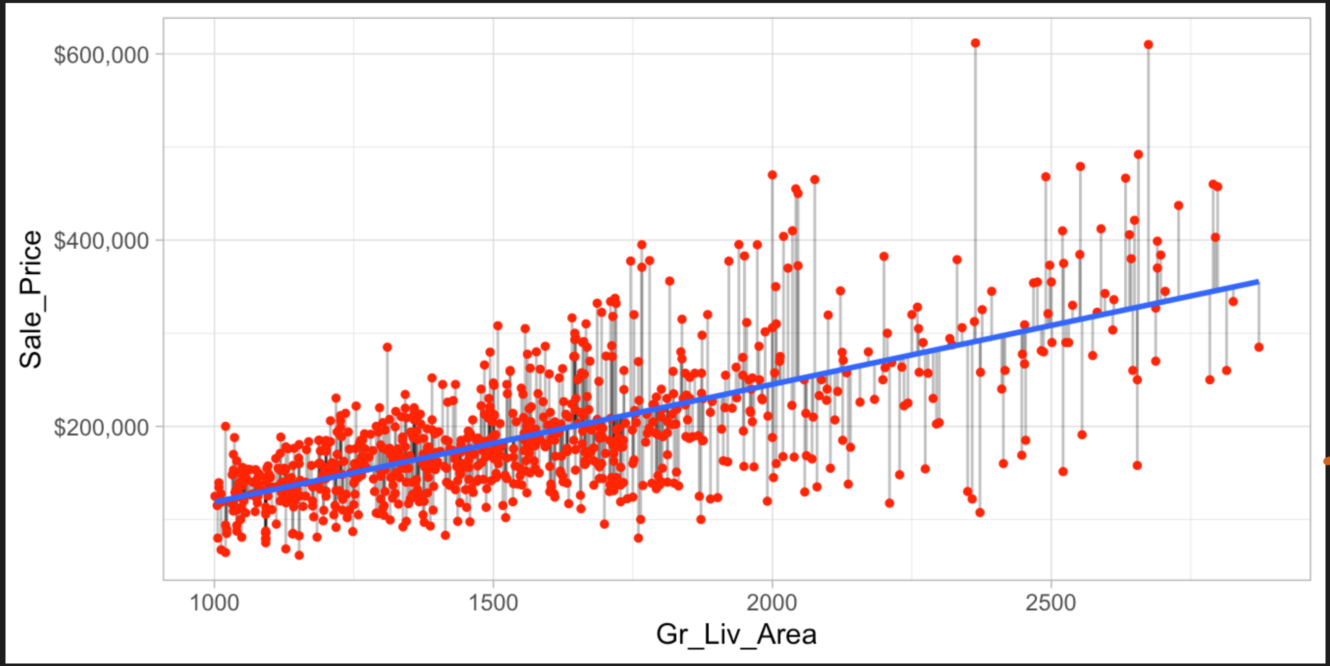
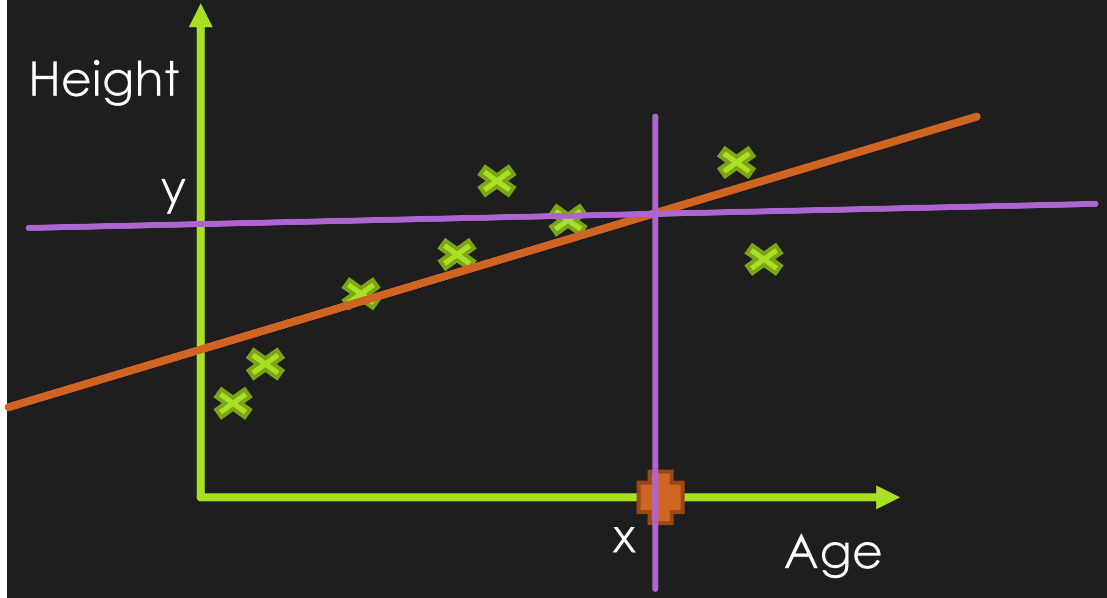
INTRODUCTION TO MACHINE LEARNING COMPSCI 4ML3

Lecture 2

Hassan Ashtiani

LINEAR CURVE-FITTING (REVIEW)





HTTPS://BRADLEYBOEHMKE.GITHUB.IO/HOML/REGULARIZED-REGRESSION.HTML

ORDINARY LEAST SQUARES (1 DIMENSION)

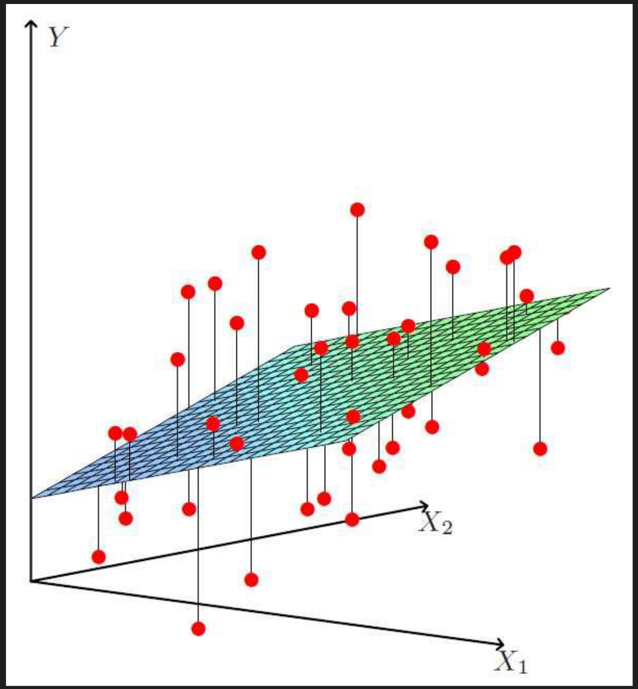
$$\{(x^i, y^i)\}_{i=1}^{n}, x^i \in \mathbb{R}, y^i \in \mathbb{R}$$

$$\underset{a,b}{\text{MIN}} \sum_{i=1}^{n} (ax^i + b - y^i)^2$$

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{COV(x, y)}{Var(x)}, b = \bar{y} - a\bar{x}$$

ORDINARY LEAST SQUARES (D DIMENSIONS)

- ASSUME $x \in \mathbb{R}^d$, $y \in \mathbb{R}$
- INSTEAD OF A LINE,
  WE NEED TO FIT A HYPERPLANE!



- HYPERPLANE EQUATION:
- $\hat{y} = w_0 + \sum_{j=1}^{d} w_j x_j = w_0 + w_1 x_1 + w_2 x_2 \ldots + w_d x_d$
- w_0 - THE y-INTERCEPT (THE BIAS)

HTTP://WWW.CS.CORNELL.EDU/COURSES/CS4758/2013SP/MATERIALS/CS4758-LIN-REGRESSION.PDF

EXAMPLE

- Estimate the price of oil based on two properties:
- (1) price of gold and (2) world GDP
- $x \in ?$
- Input data: $\{(x^i, y^i)\}_{i=1}^{n}$
- $\hat{y} = w_0 + w_1 x_1 + w_2 x_2$
- Find w_0, w_1, w_2 that give the best estimate

ORDINARY LEAST SQUARES (D-DIMENSIONS)

- SIMPLIFICATION: **HOMOGENEOUS** HYPERPLANES
  - $w_0 = 0$
  - $\hat{y} = w_1 x_1 + w_2 x_2 \ldots + w_d x_d$
  - $\hat{y} = \langle w, x \rangle = w^T x = x^T w, \qquad w = (w_1, \ldots, w_d)$
- FIND/LEARN w_j'S FROM THE DATA

$$\min_{w_1,\ldots,w_d \in \mathbb{R}} \sum_{i=1}^{n} (\hat{y}^i - y^i)^2$$

OPTIMIZE DIRECTLY?

$$\min_{w_1,\ldots,w_d \in \mathbb{R}} \sum_{i=1}^{n} (\hat{y}^i - y^i)^2 =$$

MATRIX FORM OLS
(ORDINARY LEAST SQUARES)

- $X_{n\times d} = \begin{pmatrix} x_1^1 & \cdots & x_d^1 \\ \vdots & \ddots & \vdots \\ x_1^n & \cdots & x_d^n \end{pmatrix}, Y_{n\times 1} = \begin{pmatrix} y^1 \\ \ldots \\ y^n \end{pmatrix}, W_{d\times 1} = \begin{pmatrix} w_1 \\ \ldots \\ w_d \end{pmatrix}$

PREDICTION IN VECTOR FORM

- FIND/LEARN $W_{d\times 1}$ FROM THE DATA (SOON)
- GIVEN x AND W $= W_{d\times 1}$, WHAT SHOULD $\hat{y}$ BE?
- $\hat{y} =$

FINDING W

- OBJECTIVE: $\sum_{i=1}^{n}\left(\widehat{y^i} - y^i\right)^2 = \sum_{i=1}^{n}(\langle w, x^i\rangle - y^i)^2$
- DEFINE
- $\Delta = \begin{pmatrix} \Delta_1 \\ \dots \\ \Delta_n \end{pmatrix} = \begin{pmatrix} x_1^1 & \cdots & x_d^1 \\ \vdots & \ddots & \vdots \\ x_1^n & \cdots & x_d^n \end{pmatrix} \begin{pmatrix} w_1 \\ \dots \\ w_d \end{pmatrix} - \begin{pmatrix} y^1 \\ \dots \\ y^n \end{pmatrix} = \begin{pmatrix} \widehat{y^1} - y^1 \\ \dots \\ \widehat{y^n} - y^n \end{pmatrix}$

FINDING W

- $\Delta = \begin{pmatrix} \Delta_1 \\ \dots \\ \Delta_n \end{pmatrix} = \begin{pmatrix} x_1^1 & \cdots & x_d^1 \\ \vdots & \ddots & \vdots \\ x_1^n & \cdots & x_d^n \end{pmatrix} \begin{pmatrix} w_1 \\ \dots \\ w_d \end{pmatrix} - \begin{pmatrix} y^1 \\ \dots \\ y^n \end{pmatrix}$
- OBJECTIVE FUNCTION: $\sum_{i=1}^{n} (\Delta_i)^2$
- $\min_{W \in \mathbb{R}^{d\times 1}} \sum_{i=1}^{n} (\Delta_i)^2 = \min_{W \in \mathbb{R}^{d\times 1}} \langle \Delta, \Delta \rangle = \min_{W \in \mathbb{R}^{d\times 1}} \|\Delta\|_2^2 =$

$$\min_{W \in \mathbb{R}^{d\times 1}} \|XW - Y\|_2^2$$

OLS SOLUTION

$$W^{LS} = (X^T X)^{-1} X^T Y$$

- Verify dimensions
  - Compare to $a = \frac{cov(x,y)}{Var(x)}$ for $d = 1$
- What if $X^T X$ is not invertible?