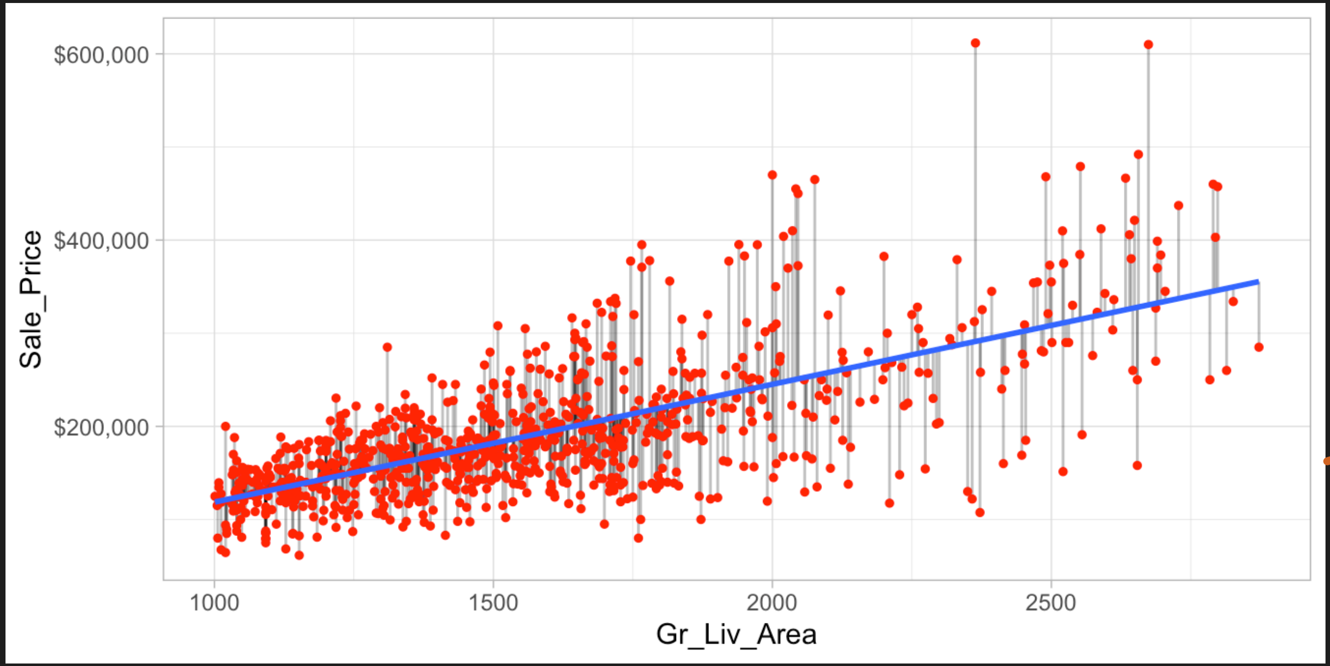
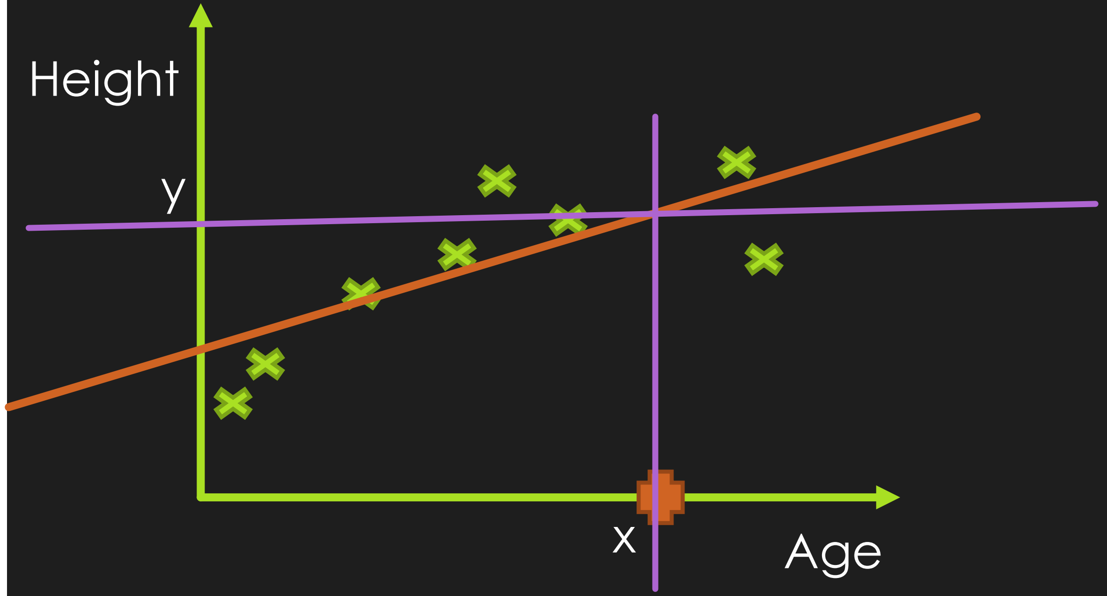
INTRODUCTION TO MACHINE LEARNING COMPSCI 4ML3

LECTURE 2

HASSAN ASHTIANI

LINEAR CURVE-FITTING (REVIEW)





[HTTPS://BRADLEYBOEHMKE.GITHUB.IO/HOML/REGULARIZED-REGRESSION.HTML](https://bradleyboehmke.github.io/HOML/regularized-regression.html)

ORDINARY LEAST SQUARES (1 DIMENSION)

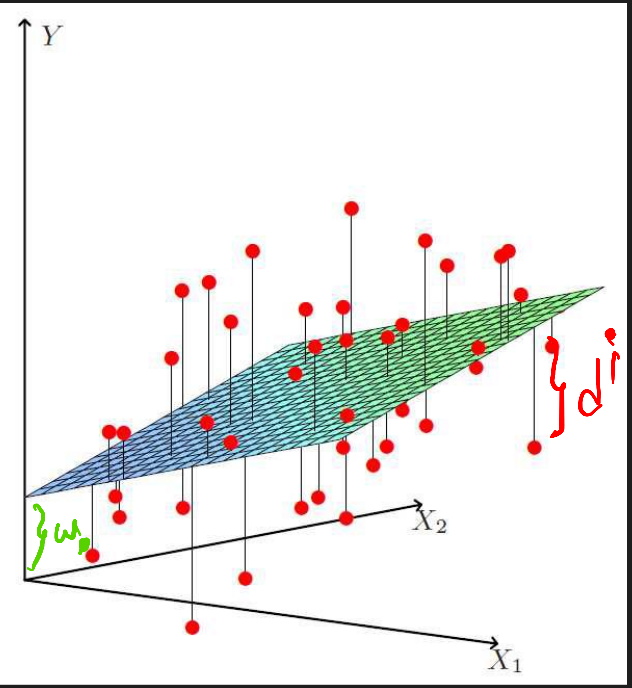
$$\left\{\left(x^i, y^i\right)\right\}_{i=1}^{n}, x^i \in \mathbb{R}, y^i \in \mathbb{R}$$

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

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

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

($x^1 = (110, 56)$, $y^1 = (70)$

$x_1 = 61$, $x_2 = (61)$

$\hat{y} = ---$

ORDINARY LEAST SQUARES (D-DIMENSIONS)

- SIMPLIFICATION: **HOMOGENEOUS** HYPERPLANES
  - $w_0 = 0$
  - $\hat{y} = w_1 x_1 + w_2 x_2 \ldots + w_d x_d = \sum_{j=1}^{d} w_j x_j$
  - $\hat{y} = \langle w, x \rangle = w^T x = x^T w,$ $\quad \underset{d \times 1}{w} = (w_1, \ldots, w_d)^T$
- 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?

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

$$0 = \frac{\partial}{\partial w_1} = 2 \sum (w^T x^i - y^i) x^i_1 = 2 \sum_{i=1}^{n} \left(\sum_j x^i_j w_j - y^i \right) x^i_1$$

$$0 = \frac{\partial}{\partial w_2} =$$

$$\vdots$$

$$0 = \frac{\partial}{\partial w_d} =$$

d equations

d unknowns

MATRIX FORM OLS (ORDINARY LEAST SQUARES)

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

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?
- $\hat{y} =$

$$\hat{y} = w^T x = x^T w = \sum_{j=1}^{d} w_j x_j = \langle w, x \rangle$$

FINDING W

$\sum \Delta_i^2$

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

$\Delta = X W - Y$

first row

$\sum x_j^1 w_j = \underbrace{x^{1^T} w}_{= y^1}$

FINDING W

$\|v\|_2 = \sqrt{\sum v_i^2}$

$\|v\|_p = \left(\sum |x_i|^p\right)^{1/p}$

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

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

$XW - Y = \Delta$

$\Updownarrow$

$\|XW - Y\|_2^2 = \|\Delta\|_2^2$

$X_{n\times d}$, $X^T : d\times n$

OLS SOLUTION

$$W^{LS}_{d\times 1} = \underbrace{\underbrace{\underbrace{(\underset{d\times n}{X^T}\underset{n\times d}{X})^{-1}}_{d\times d}\underset{d\times n}{X^T}}_{d\times n} Y_{n\times 1}}_{d\times 1}$$

- VERIFY DIMENSIONS
  - COMPARE TO $a = \frac{cov(x,y)}{Var(x)}$ FOR $d = 1$
- WHAT IF $X^T X$ IS NOT INVERTIBLE?

$$(x^i, y^i) = (0, 0)$$