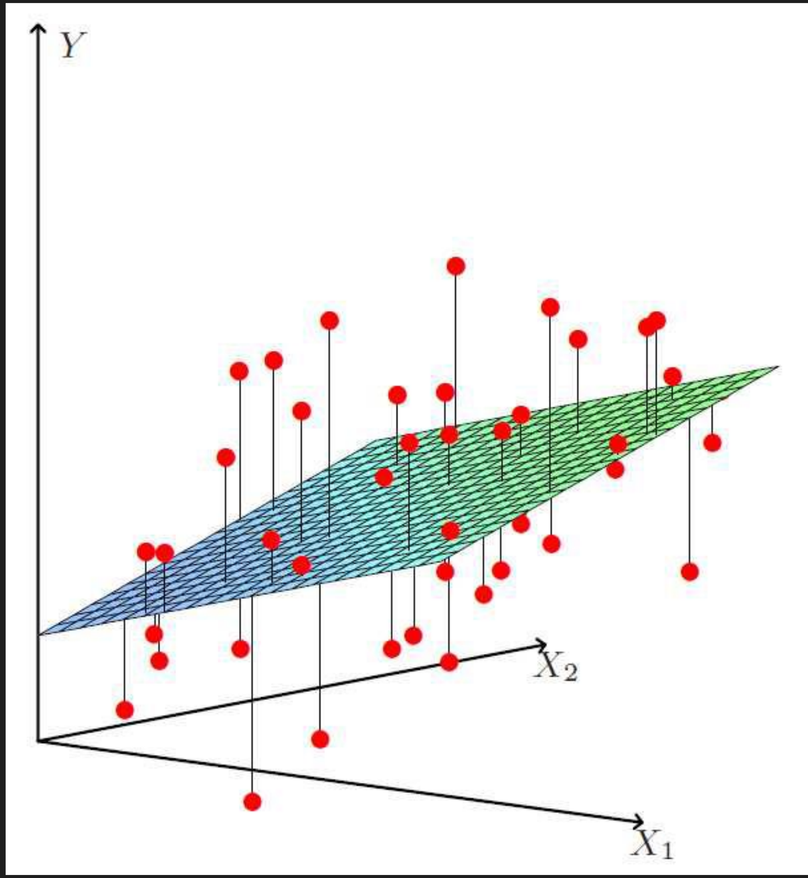
INTRODUCTION TO MACHINE LEARNING COMPSCI 4ML3

LECTURE 3

HASSAN ASHTIANI

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!
- WHY ARE THE LINES VERTICAL?
  - ANY DIFFERENT IF WE MINIMIZE THE DISTANCE TO THE HYPERPLANE?



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

MATRIX FORM OLS

- $\Delta = \begin{pmatrix} \Delta_1 \\ \dots \\ \Delta_n \end{pmatrix} = \begin{pmatrix} x_1^1 & \cdots & x_d^1 \\ \vdots & \ddots & \vdots \\ x_1^n & \cdots & x_d^n \end{pmatrix} \begin{pmatrix} w_1 \\ \dots \\ w_d \end{pmatrix} - \begin{pmatrix} y^1 \\ \dots \\ y^n \end{pmatrix}$

$$\underset{W \in \mathbb{R}^{d \times 1}}{\text{MIN}} \sum_{i=1}^{n} (\Delta_i)^2 = \underset{W \in \mathbb{R}^{d \times 1}}{\text{MIN}} < \Delta, \Delta > = \underset{W \in \mathbb{R}^{d \times 1}}{\text{MIN}} \|\Delta\|_2^2$$

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

TAKING THE "DERIVATIVE"

- REAL-VALUED FUNCTION OF A VECTOR

GRADIENT:

- VECTOR-VALUED FUNCTION OF A VECTOR

JACOBIAN:

MATRIX/VECTOR CALCULUS

- $u, v \in R^n$
- $g(u) = u^T v$

- $\nabla u(g)=$

MATRIX/VECTOR CALCULUS

- $A \in R^{m \times n}, u \in R^{n}$
- $g(u) = Au$

- $\nabla u(g)=$

MATRIX/VECTOR CALCULUS

- $A \in R^{m \times n}, u \in R^n$
- $g(u) = u^T A\, u$

- $\nabla u(g)=$

SOLVING OLS

$f(W) = \min_{W \in \mathbb{R}^{d \times 1}} \|XW - Y\|_2^2.$ WHAT IS ∇f?

SOLVING OLS

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

- Degenerate case when $X^T X$ is not invertible?

BIAS/INTERCEPT TERM

- We are missing the bias term (w_0)

$$\min_{w_0, w_1, \ldots, w_d \in \mathbb{R}} \sum_{i=1}^{n} \left(w_1 x_1^i + \cdots + w_d x_d^i + w_0 - y^i\right)^2$$

- Matrix form with the bias term?

$$\min_{W \in \mathbb{R}^{d \times 1}, w_0 \in \mathbb{R}} \left\| XW + \begin{pmatrix} w_0 \\ w_0 \\ \ldots \\ w_0 \end{pmatrix} - Y \right\|_2^2$$

EXAMPLE

BIAS/INTERCEPT TERM

- ADD A NEW AUXILIARY DIMENSION TO THE DATA
- $X'_{n\times(d+1)} = \begin{pmatrix} x_1^1 & \cdots & x_d^1 & 1 \\ \vdots & \ddots & \vdots & 1 \\ x_1^n & \cdots & x_d^n & 1 \end{pmatrix}, W'_{(d+1)\times 1} = \begin{pmatrix} w_1 \\ \dots \\ w_d \\ w_0 \end{pmatrix}$
- SOLVE OLS: $\min\limits_{W' \in \mathbb{R}^{(D+1)\times 1}} \|X'W' - Y\|_2^2$
- w_0 WILL BE THE BIAS TERM!

SOME EXAMPLES

- OLS NOTEBOOK