# CS456: Machine Learning

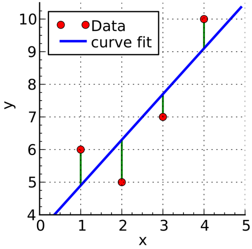
## Linear Least Square Regression Model

Jakramate Bootkrajang

Department of Computer Science

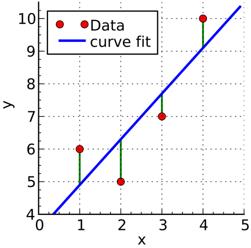
Chiang Mai University

# Linear Regression Problem




- Recall that in linear regression problem we are given a set of data $(\mathsf{x}_i, y_i)$ , $i = 0, \ldots, N$ (red dots)
- And we want to find a linear function $f(\mathsf{x})$ that best describes the data

# Motivation




- In most of the cases, $f(\mathbf{x})$ cannot perfectly describe the data
- There is almost always mismatch between $y_i$ (the true value) and the value given by the model $f(\mathbf{x}_i)$
- The manitude of the mismatch is length of green line

# Sum of squared errors

- The error of the estimation of each $y_i$ is given by

$$\epsilon_i = y_i - f(\mathsf{x}_i) \tag{1}$$
$$= y_i - a\mathsf{x}_i \tag{2}$$

- Since $\epsilon_i$ can be either positive or negative, their values might cancel out

- So we square the error terms and get

$$\epsilon_i = (y_i - a\mathsf{x}_i)^2 \tag{3}$$

- Summing all the error terms we have

$$\epsilon = \sum_{i=0}^{N} (y_i - a\mathsf{x}_i)^2 \tag{4}$$

# In matrix form

- We can write the sum of squared errors in matrix form

$$\epsilon = ||X\mathbf{a} - Y||_2^2 \quad (5)$$

- $X$ is a data matrix $N$ by $M$ ($N$=number of data points $M$= data dimension)
- $Y$ is matrix of responses
- $\mathbf{a}$ is a coefficient vector of the linear model

# In matrix form: Example

- Given $\{(3,14),(2,13),(5,19)\}$

$$X = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}, \mathsf{a} = \begin{bmatrix} a_1 \end{bmatrix}, Y = \begin{bmatrix} 14 \\ 13 \\ 19 \end{bmatrix} \tag{6}$$

- Or another example (2 dimensions)

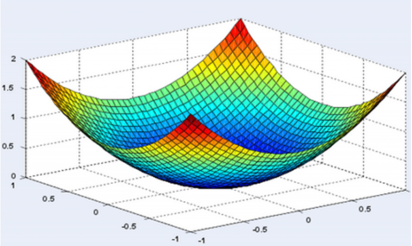
$$X = \begin{bmatrix} 3 & 2 \\ 2 & 1 \\ 5 & 4 \end{bmatrix}, \mathsf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, Y = \begin{bmatrix} 3.3 \\ 1.5 \\ 7.9 \end{bmatrix} \tag{7}$$

## Linear Least Square

- Goal: Find a vector **a** which when multiplied to all the data gives the smallest sum of squared errors (the objective function)

$$L = \min_{\mathbf{a}} ||X\mathbf{a} - Y||_2^2 \tag{8}$$

- The function in Eq.(**??**) is a convex function with a global minimum



# Where the best **a** is ?

- A point where the minimum is, is the stationary point (slope/derivative) is zero
- First we find the derivative of the objective function w.r.t **a**

$$\frac{\partial L}{\partial \mathsf{a}} = \frac{\partial ||X\mathsf{a} - Y||_2^2}{\partial \mathsf{a}} \tag{9}$$

$$= 2(X\mathsf{a} - Y)X \tag{10}$$

- Next find **a** which gives zero slope

$$(X\mathsf{a} - Y)X = 0 \tag{11}$$

$$X^T X\mathsf{a} - X^T Y = 0 \tag{12}$$

$$\mathsf{a} = (X^T X)^{-1} X^T Y \tag{13}$$

# Implementation example

Let's switch to Google's colab – https://colab.research.google.com/drive/1o94L5SKWkb4f8L08q92FthRc2Qc8YpVJ