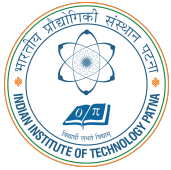
Introduction to Data Science

Linear Regression



Arijit Mondal

Dept. of Computer Science & Engineering

Indian Institute of Technology Patna

arijit@iitp.ac.in

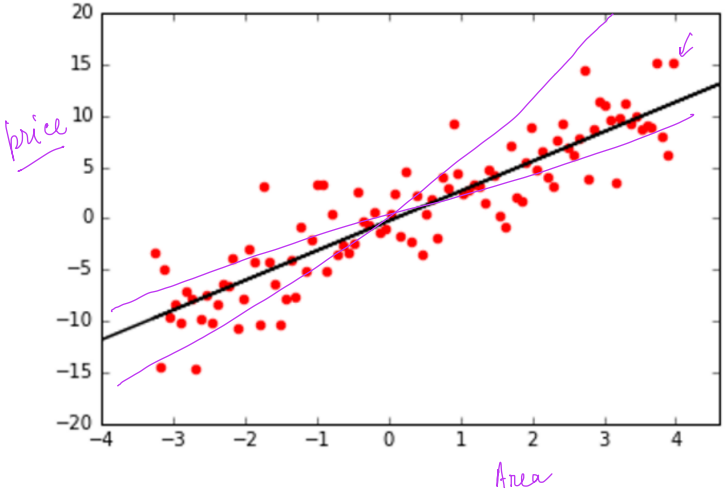
Introduction

- Most representative machine learning method
- Easy to understand
- Generally appropriate as a default / baseline model
  - Price of a house grows linearly with area
  - Weight increases linearly with the food eaten

Introduction

- Most representative machine learning method
- Easy to understand
- Generally appropriate as a default / baseline model
  - Price of a house grows linearly with area
  - Weight increases linearly with the food eaten
- Statistician's rule - if you want a linear function, just measure two points only!!! ✓✓

Example



n points (x_i, y_i)

we want to fit a curve $y = f(x)$

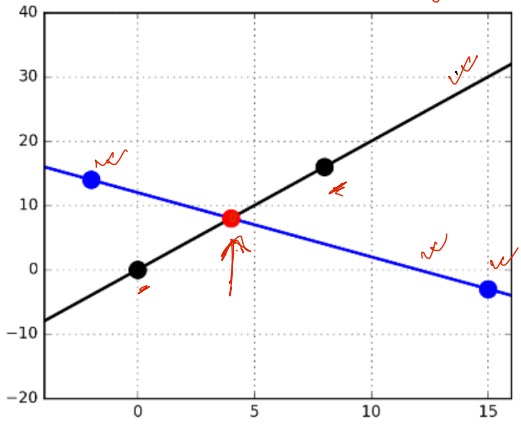
$y = w_0 + w_1 x$

$x_j \rightarrow y_j = w_0 + w_1 x_j$

image source: Data Science Design Manual

Linear regression and duality

- For solving linear systems, we find the single point that lies on n given lines
- In regression, we are given n points and we seek the line that lies on all the points
- By duality transformation both are the same



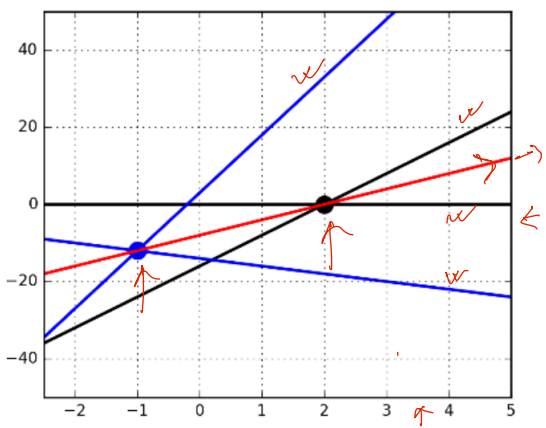
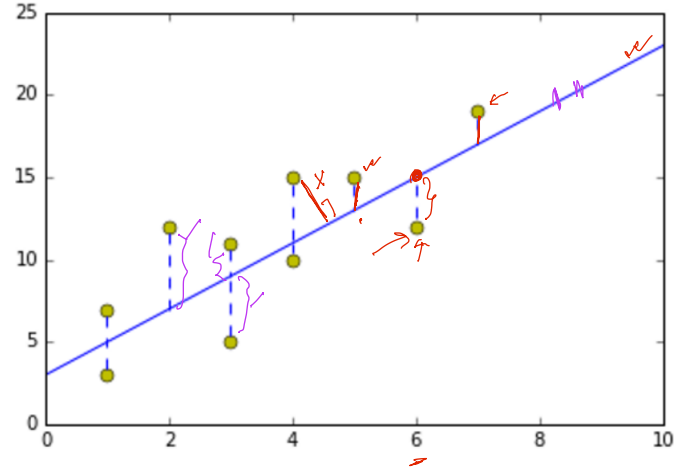


image source: Data Science Design Manual

Error in linear regression

- Residual error → $y = w_0 + w_1 x$ $x_i \rightarrow \hat{y}_i = w_0 + w_1 x_i$ $(y_i - \hat{y}_i)$
- Least squares regression minimizes the sum of the squares of the residuals of all points



$L_i = (y_i - \hat{y}_i)$ → +ve, 0, -ve

$\sum (L_i)^2$ → +ve

image source: Data Science Design Manual

Finding the optimal fit

$\|x\| = x^T x$

$E = \sum (y_i - \hat{y}_i)^2 \rightarrow$ minimize $\leftarrow$

$\hat{y}_i \rightarrow L(w_0, w_1)$

$$E = \sum (y_i - w_0 - w_1 x_i)^2$$

$$\frac{\partial E}{\partial w_0} = 0 = \sum (y_i - w_0 - w_1 x_i) \quad (1)$$

$$= \sum y_i - w_0 n - w_1 \sum x_i$$

$$\Rightarrow \bar{y} - w_0 - w_1 \bar{x} = 0$$

$$\frac{\partial E}{\partial w_1} = 0 = \sum (y_i - w_0 - w_1 x_i) x_i$$

$$= \sum x_i y_i - \sum x_i w_0 - \sum w_1 x_i^2$$

$$= \sum x_i y_i - \sum x_i (\bar{y} - w_1 \bar{x}) - \sum w_1 x_i^2$$

$$= \sum x_i y_i - \bar{y} \cdot n \bar{x} + w_1 \bar{x}^2 - w_1 \sum x_i^2$$

$$\sum (x_i - \bar{x})(y_i - \bar{y}) + w_1 \sum (x_i - \bar{x})^2 = 0$$

$$w_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\rightarrow y = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_{m-1} x_{m-1}$$

$$y_{n\times 1} = [A]_{n\times m} \cdot [w]_{m\times 1}$$

$$(y_{m\times 1} - \hat{y}_{n\times 1}) = (y - Aw)$$

$\partial E/\partial w_0, \frac{\partial E}{\partial w_1}, \ldots$

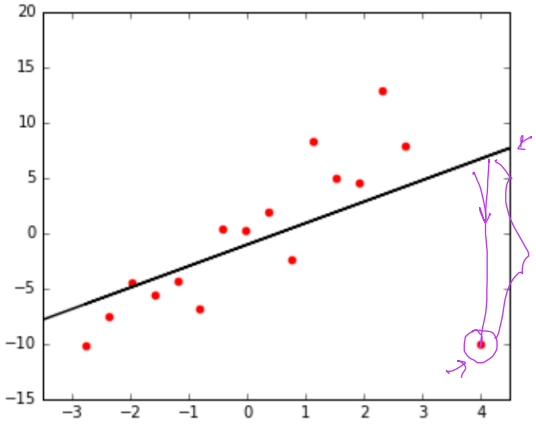
$$E = (y - Aw)^T (y - Aw) = y^T y - y^T Aw - w^T A^T y + w^T A^T A w$$

$$\frac{\partial E}{\partial w} = -2A^T y + 2A^T A w = 0$$

$$w = (A^T A)^{-1} A^T y$$

pinv

Better models: Removing outliers



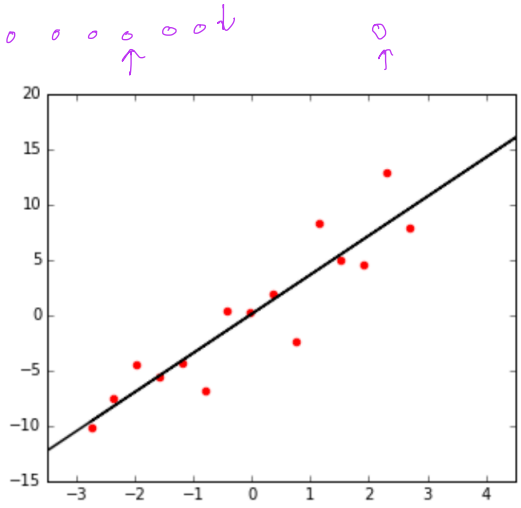
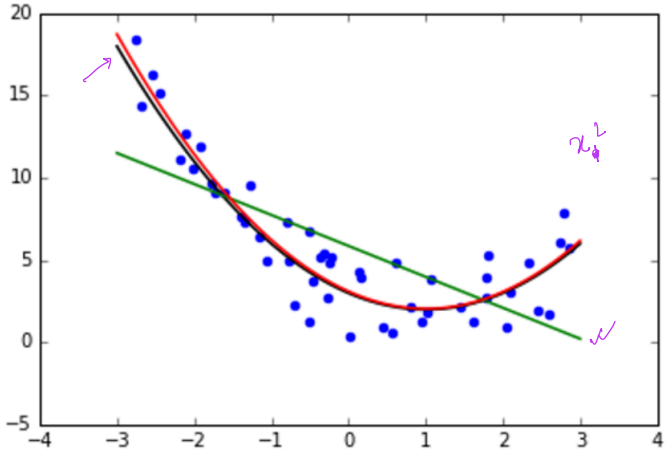


image source: Data Science Design Manual

Better models: Non-linear functions



$y = w_0 + w_1 x$

$y = w_0 + w_1 x_1 + w_2 x_2 + \cdots$

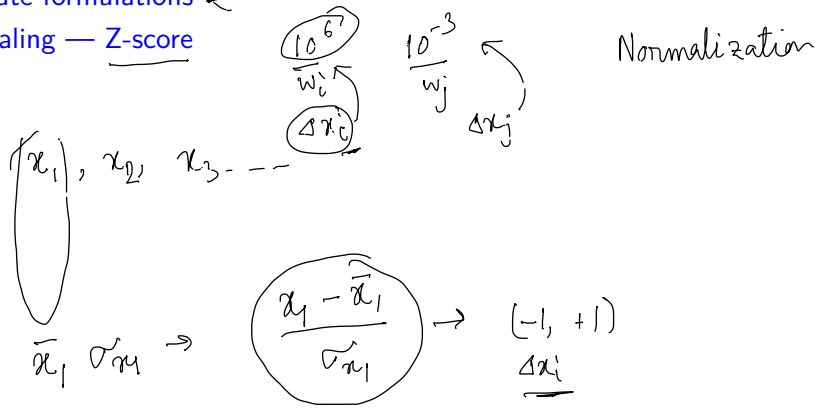
$x_2 = x_1^2$

$x_3 = x_1 x_2$

image source: Data Science Design Manual

Feature and target scaling

- Unreadable coefficient
- Numerical imprecision
- Inappropriate formulations
- Feature scaling — Z-score



Regression as parameter fitting

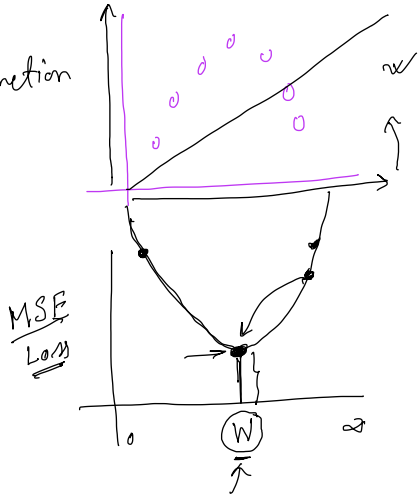
$y = w_0 + w_1 x$

$y = wx$

$0 - \infty$

$w \rightarrow MSE \rightarrow$ cost function

Loss



Convex parameter spaces

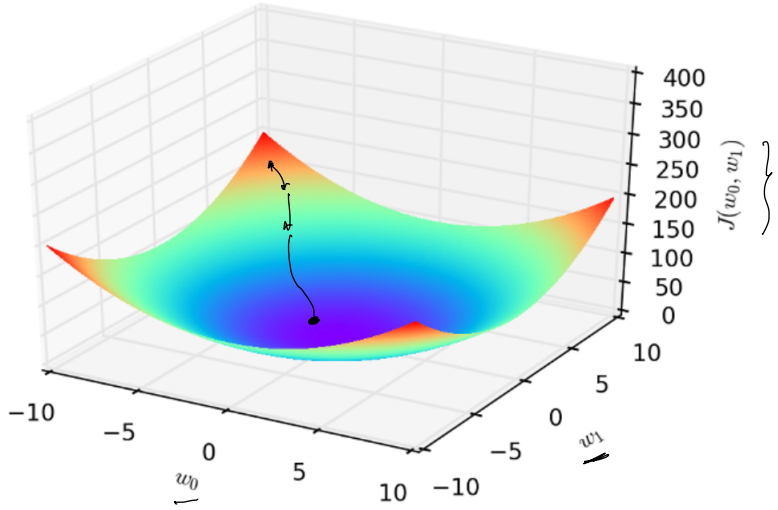


image source: Data Science Design Manual