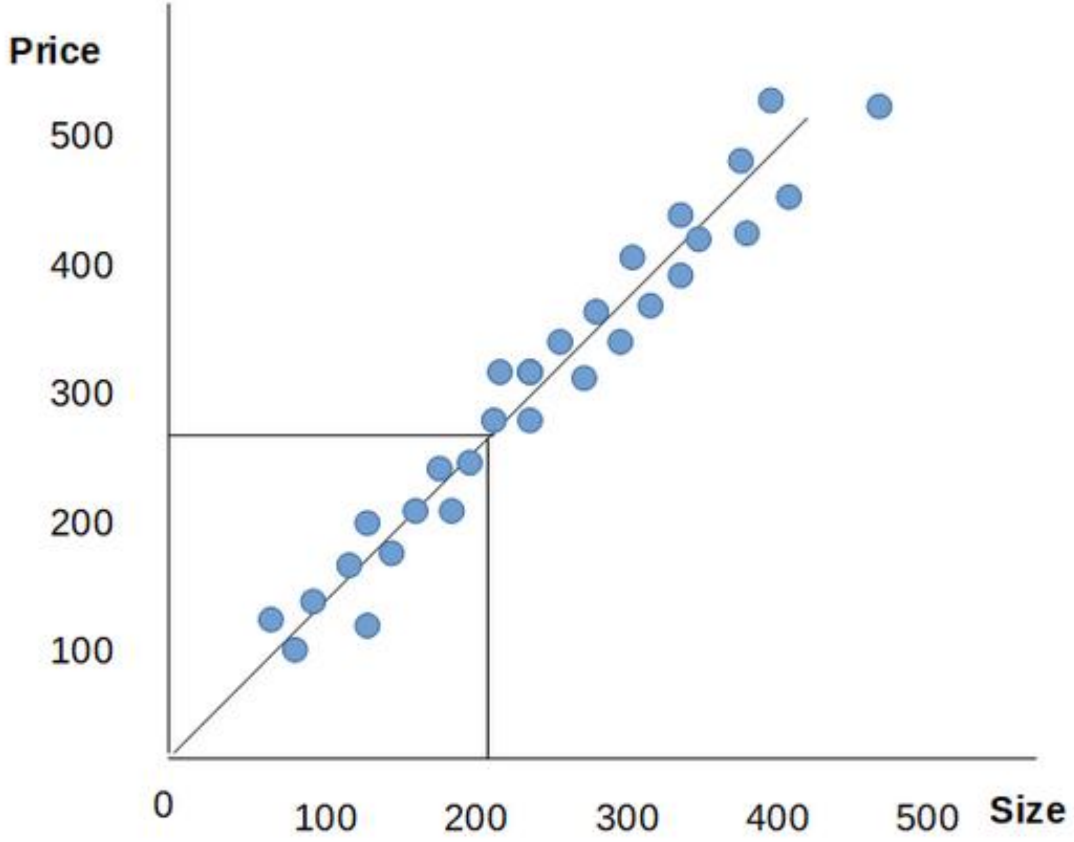


Least Square Regression

PhD. Msc. David C. Balderas S.

Least-Square Regression

What is regression?

Given n data points $S_r = (x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)$, best fit $y = f(x)$ to the data. The best fit is generally based on minimizing the sum of the square of the residuals,

Residual at a point is

$$\varepsilon_i = y_i - f(x_i)$$

The sum of the square of the residuals

$$S_r = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

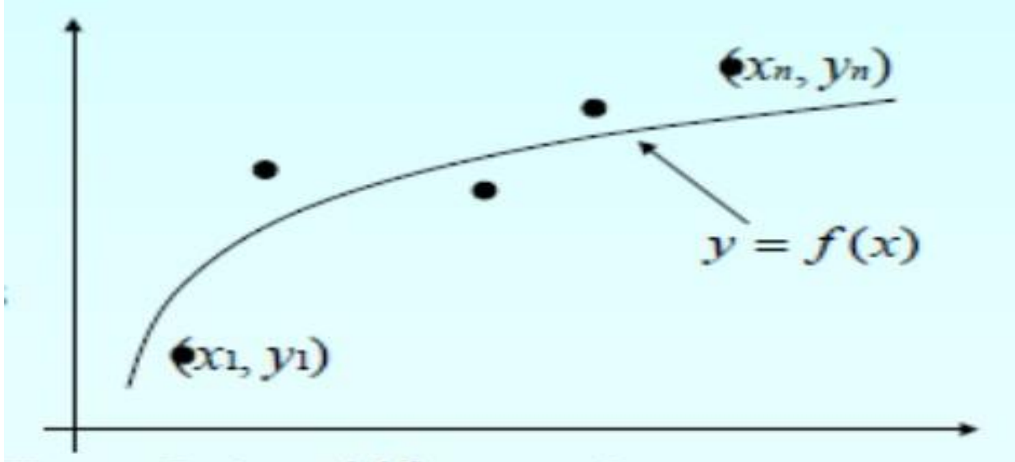


Figure. Basic model for regression

Least-Square Regression

- What is regression?
  - Is a tool for the investigator of relationships between variables
- Types of regression:
  - Linear Regression
  - Nonlinear Regression
    - Polynomial Regression
    - Exponential Regression Etc...

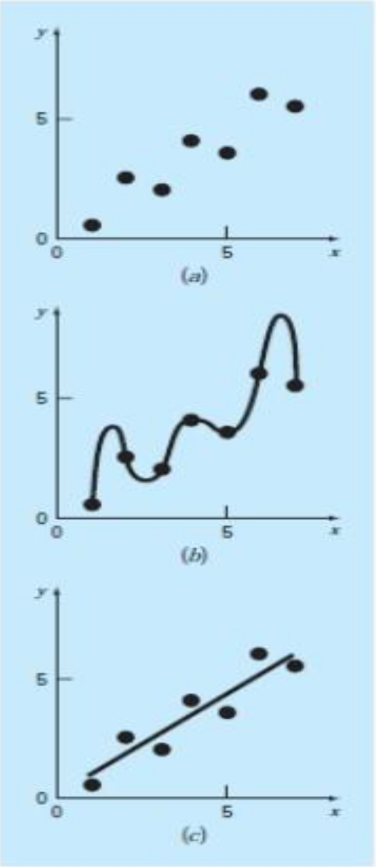


FIGURE 17.1
(a) Data exhibiting significant error. (b) Polynomial fit oscillating beyond the range of the data. (c) More satisfactory result using the least-squares fit.

Linear Regression

Given n data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ best fit to the data $y = a_0 + a_1 x + \varepsilon$

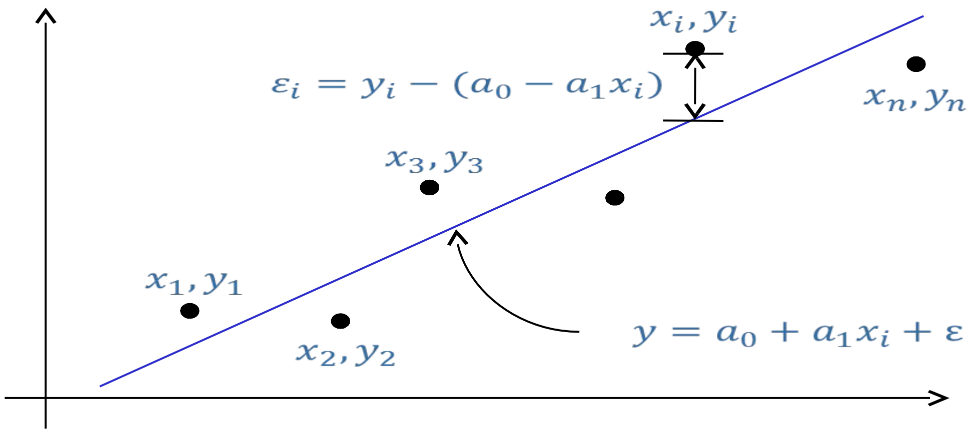


Figure. Linear regression of y vs. x data showing residuals at a typical point,

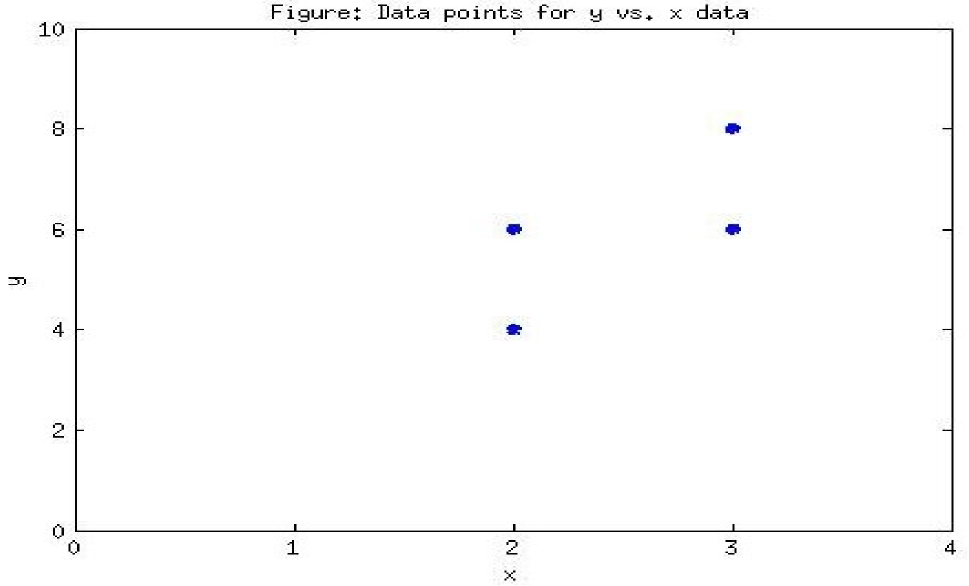
Does minimizing $S_r = \sum_{i=1}^{n} \varepsilon_i$ work as criterion, where

$$\varepsilon_i = y_i - (a_0 + a_1 x_i)$$

Example for Criterion # 1- Minimize the sum residual errors for all the available data

Given the data points (2,4), (3,6), (2,6) and (3,8), best fit the data to a straight line using Criterion # 1: $\sum_{i=1}^{n} \varepsilon_i$

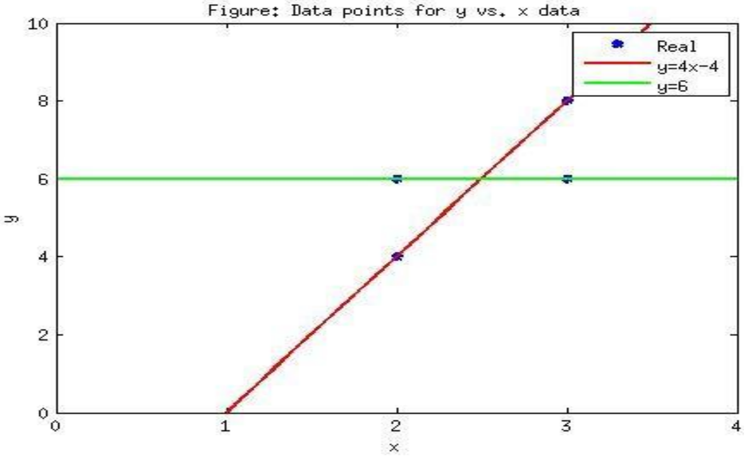
x	y
2.0	4.0
3.0	6.0
2.0	6.0
3.0	8.0



Example for Criterion # 1

Using $y_1 = 4x - 4$ or $y_2 = 6$ with $\sum_{i-1}^{n} \varepsilon_i$

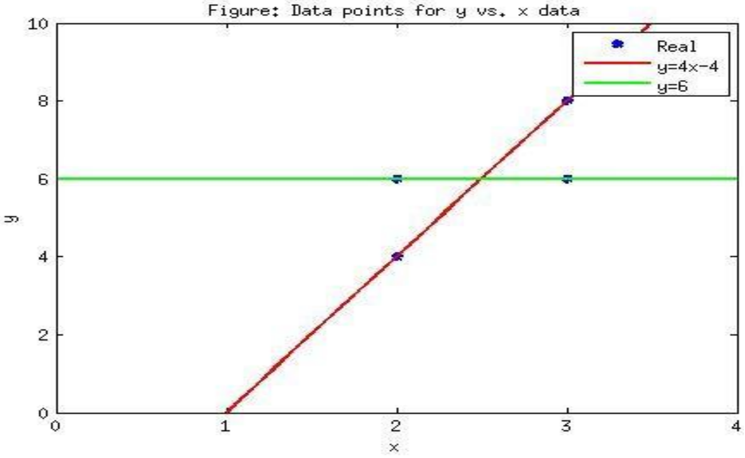
x	y	y1	y2	
2.0	4.0			
3.0	6.0			
2.0	6.0			
3.0	8.0			



Example for Criterion # 1

Using $y_1 = 4x - 4$ or $y_2 = 6$ with $\sum_{i-1}^{n} \varepsilon_i$

x	y	y1	y2		
2.0	4.0	4.0	6.0	0.0	-2.0
3.0	6.0	8.0	6.0	-2.0	0.0
2.0	6.0	4.0	6.0	2.0	0.0
3.0	8.0	8.0	6.0	0.0	2.0
				0.0	0.0



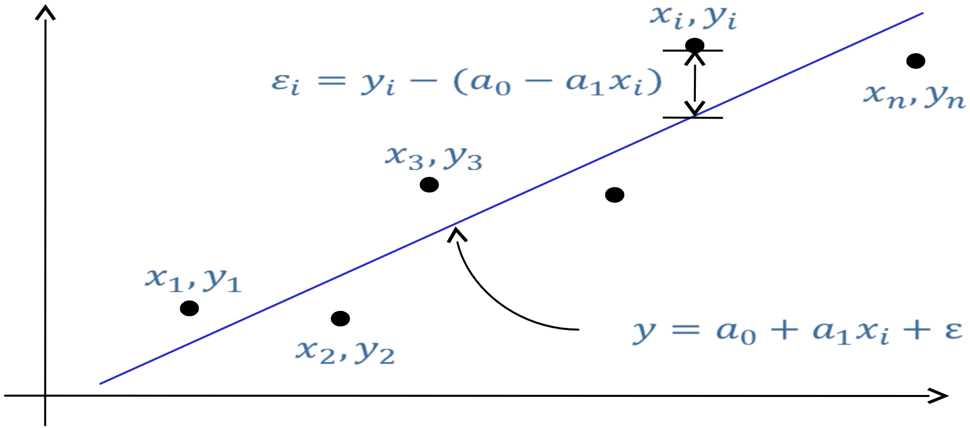
- The sum of the errors has been made as small as possible, that is 0, but the regression model is not unique.
- Hence the above criterion of minimizing the sum of the residuals is a bad criterion.

Criterion # 2 – Minimize the sum of the absolute value of the discrepancies

Does minimizing $\sum_{i-1}^{n} |\varepsilon_i|$ work as criterion, not worrying if they cancel as each other

$$\sum_{i=1}^{n} |\varepsilon_i| = \sum_{i=1}^{n} |y_i - a_0 - a_1 x_i|$$

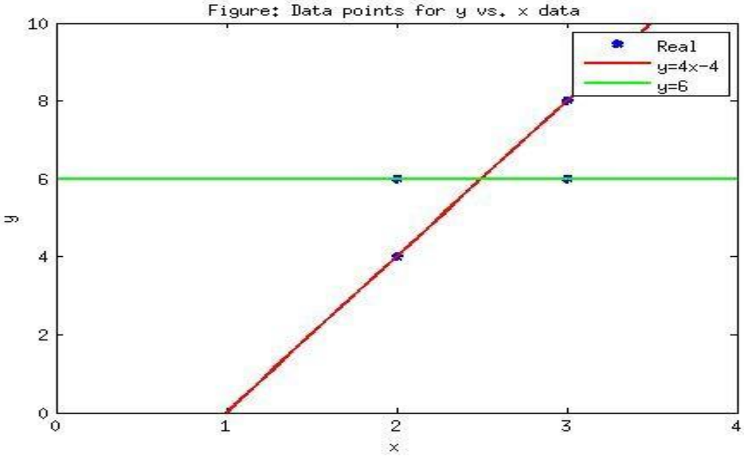
Figure. Linear regression of y vs. x data showing residuals at a typical point, x_j



Example for Criterion # 2

Using $y_1 = 4x - 4$ or $y_2 = 6$ with $\sum_{i-1}^{n} \varepsilon_i$

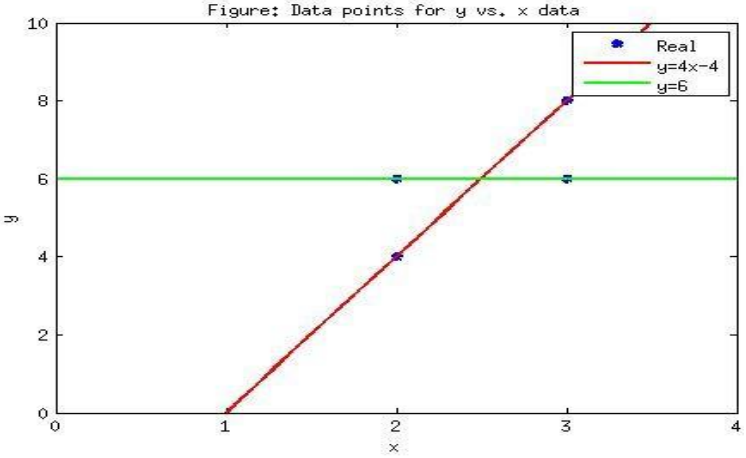
x	y	y1	y2	
2.0	4.0			
3.0	6.0			
2.0	6.0			
3.0	8.0			



Example for Criterion # 2

Using $y_1 = 4x - 4$ or $y_2 = 6$ with $\sum_{i-1}^{n} \varepsilon_i$

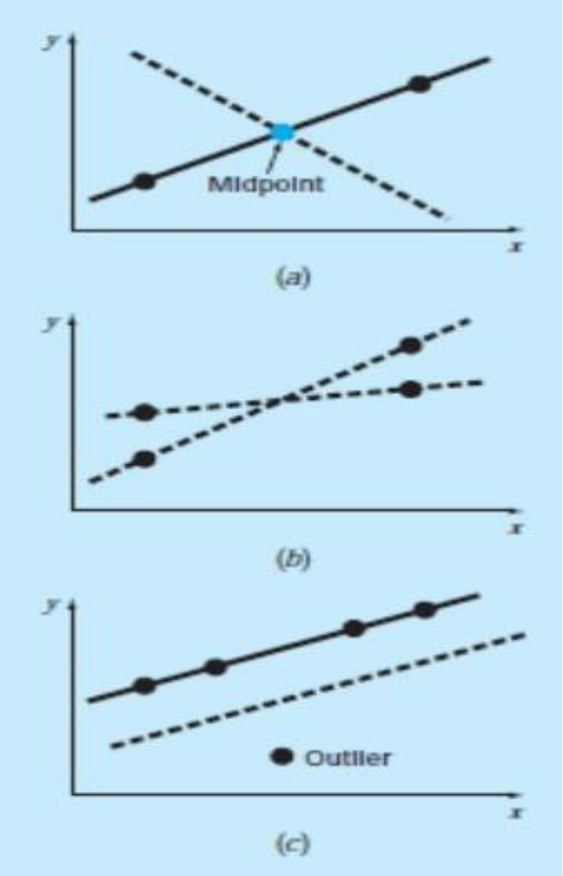
x	y	y1	y2		
2.0	4.0	4.0	6.0	0.0	2.0
3.0	6.0	8.0	6.0	2.0	0.0
2.0	6.0	4.0	6.0	2.0	0.0
3.0	8.0	8.0	6.0	0.0	2.0
				4.0	4.0



- The sum of the errors has been made as small as possible, that is 4, but the regression model is not unique.
- Hence the above criterion of minimizing the sum of the residuals is a bad criterion.

The three strategies so far!

Examples of some criteria for “best fit” that are inadequate for regression: (a) minimizes the sum of the residuals, (b) minimizes the sum of the absolute values of the residuals, and (c) minimizes the maximum error of any individual point.



Least Squares Criterion

- The least squares criterion minimizes the sum of the square of the residuals in the model, and also produces a unique line.

$$S_r = \sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

S_r :Sum of square of residuals

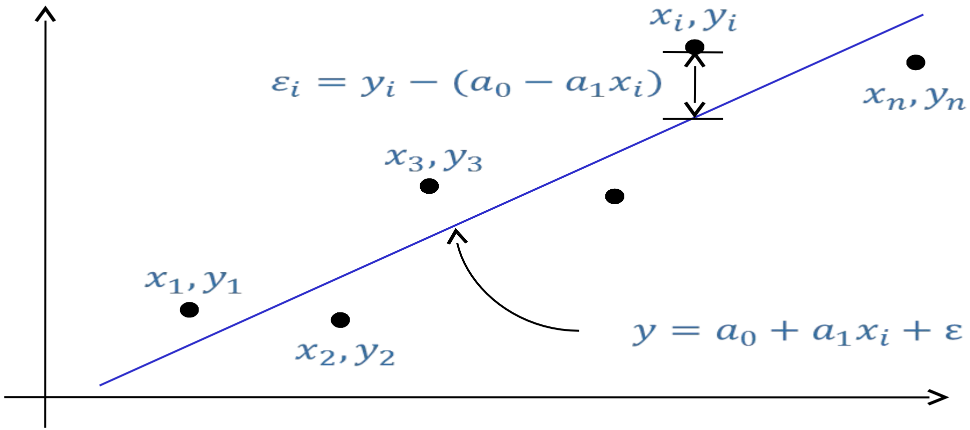


Figure. Linear regression of y vs. x data showing residuals at a typical point, x_j

Finding Constants of Linear Model

- Minimize the sum of the square of the residuals:

$$S_r = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

with respect of the constants of regression model.

- To find a_0 and a_1 we minimize S_r with respect to a_0 and a_1

$$\frac{\partial S_r}{\partial a_0} = 2\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i)(-1) = 0$$

$$\frac{\partial S_r}{\partial a_1} = 2\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i)(-x_i) = 0$$

giving

$$\sum_{i=1}^{n} a_0 + \sum_{i=1}^{n} a_1 x_i = \sum_{i=1}^{n} y_i$$

$$(a_0 = \bar{y} - a1\bar{x})$$

$$\sum_{i=1}^{n} a_0 + \sum_{i=1}^{n} a_1 x_i = \sum_{i=1}^{n} y_i x_i$$

Finding Constants of Linear Model

Solving for a_0 and a_1

$$a_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2}$$

and

$$a_0 = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2}$$

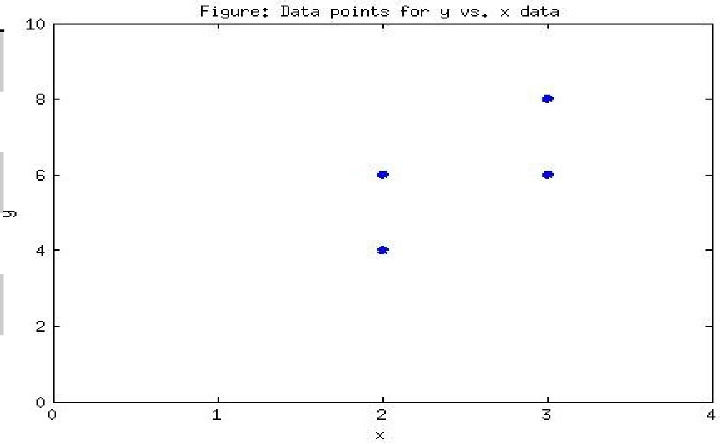
$$a_0 = \bar{y} - a_1 \bar{x}$$

Example 1

- Linear Regression

$$S_r = \sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

x	y	
2.0	4.0	
3.0	6.0	
2.0	6.0	
3.0	8.0	



Example 1

- Linear Regression

$$y = a_0 + a_1 x$$

	2.0	4.0	4.0	8.0
	3.0	6.0	9.0	18.0
	2.0	3.0	4.0	12.0
	3.0	8.0	9.0	24.0
$\sum_{i=1}^{n}$	10.0	24.0	26.0	62.0

$$a_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$a_1 = \frac{4(62) - 10(24)}{4(26) - (10)^2} = 2$$

$$a_0 = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$a_0 = \frac{26(24) - 10(62)}{4(26) - (10)^2} = 1$$

alternative method

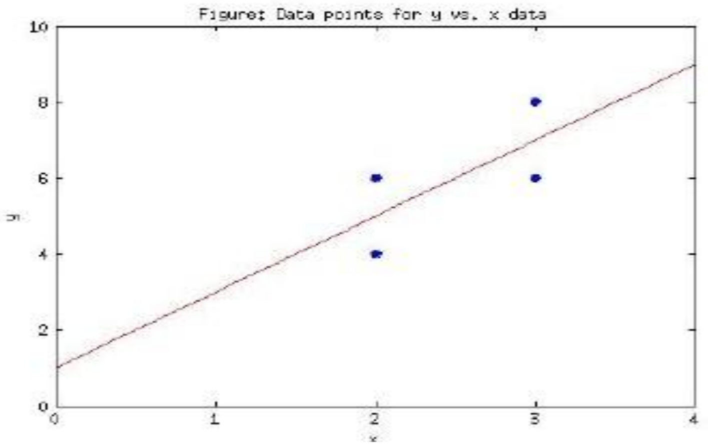
$$a_{0_2} = \bar{y} - a_1 \bar{x} = \frac{24}{4} - 2 * (\frac{10}{4}) = 1$$

Example 1

- Linear Regression $$S_r = \sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

- Obtained $y - 1 + 2x$

x	y		
2.0	4.0	5.0	-1.0
3.0	6.0	7.0	-1.0
2.0	6.0	5.0	1.0
3.0	8.0	7.0	1.0
			4

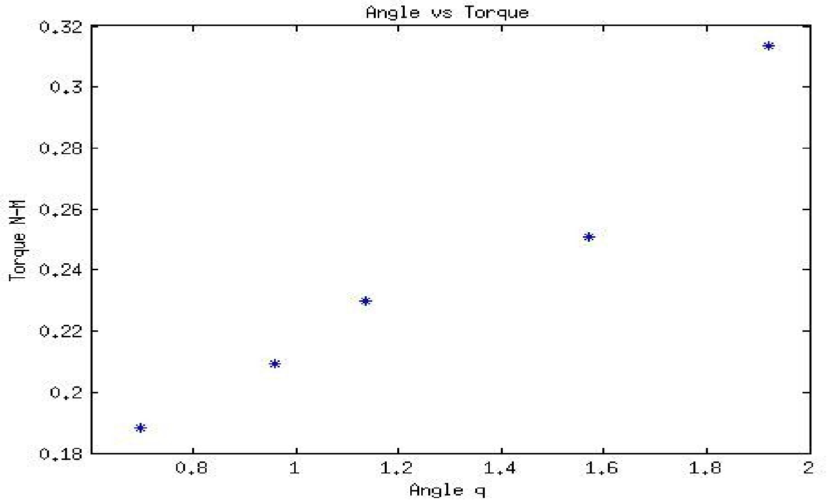


Example 2

- The torque T, needed to turn the torsion spring of a mousetrap through an angle, is given below. Find the constants for the model given by θ

$$T = k_1 + k_2\theta$$

Angle Radians	Torque, T N-m
0.698132	0.188224
0.959931	0.209138
1.134464	0.230052
1.570796	0.250965
1.919862	0.313707



Example 2

- The torque T, needed to turn the torsion spring of a mousetrap through an angle, is given below. Find the constants for the model given by

$$T = k_1 + k_2\theta$$

Angle θ	Torque, T	θ^2	$T\theta$
Radians	N-m	Radians²	N-m-Radians
0.698132	0.188224	0.487388	0.131405
0.959931	0.209138	0.921468	0.200758
1.134464	0.230052	1.2870	0.260986
1.570796	0.250965	2.4674	0.394215
1.919862	0.313707	3.6859	0.602274
$\sum_{i=1}^{5}$ 6.2831	1.1921	8.8491	1.5896

$$k_2 = \frac{n\sum_{i=1}^{5}\theta_i T_i - \sum_{i=1}^{5}\theta_i \sum_{i=1}^{n} T_i}{n\sum_{i=1}^{5}\theta_i^2 - \left(\sum_{i=1}^{5}\theta_i\right)^2}$$

$$k_2 = \frac{5(1.5896) - (6.2831)(1.1921)}{5(8.8491) - (6.2831)^2}$$

$$k_2 = 0.962461 \times 10^{-1} N - m$$

Use the average torque and angle for k1

$$\bar{T} = \frac{\sum_{i=1}^{5} T_i}{n} = \frac{1.1921}{5} = 2.3842 \times 10^{-1}$$

$$\bar{\theta} = \frac{\sum_{i=1}^{5} \theta_i}{n} = \frac{6.2731}{5} = 1.2566$$

$$k_1 = \bar{T} - k_2\bar{\theta}$$

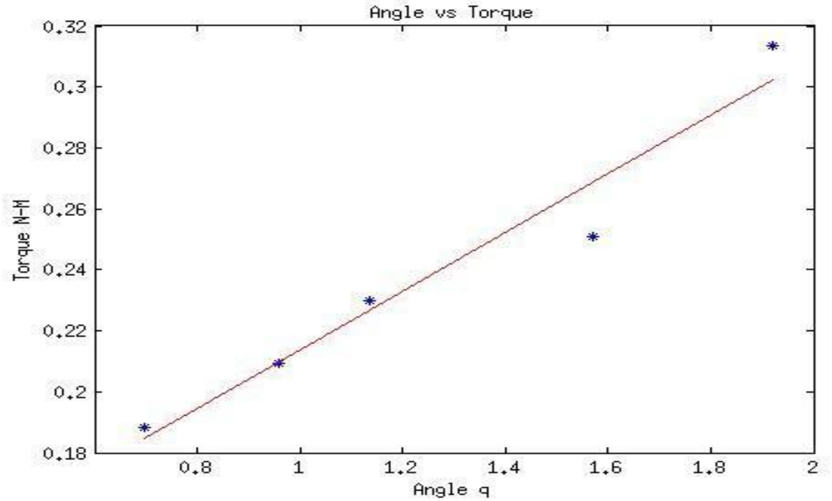
$$k_1 = 1.1767 \times 10^{-1} N - m$$

$$T = 1.1764 \times 10^{-1} + 0.9624 \times 10^{-1}\theta$$

Example 2 Results

Using linear regression, a trend line is found from the data

$$T = 1.1764 \times 10^{-1} + 0.9624 \times 10^{-1}\theta$$



Linear Regression

- When a learning model fits data by a line, we say it is linear.
- Regression is relationship between correlated variables
  - When you have a +/- change in a variable that affects the other variable +/-
- When you have just one input value (feature), linear regression is the problem if fitting a line to the data and using this line to make new predictions.
- By convention in machine learning we write: $y' = b + wx$. So instead of m, we use w.

Linear Regression with Many Features

- Often you have multiple features that can be used to predict a value.
  - For example suppose you want to predict the rent for an apartment.
  - Two relevant features are: the number of bedrooms and the apt condition from 1 (poor) to 5 (excellent)
- In this case the input x becomes a pair (x_1 , x_2) where x_1 is the number of bedrooms and x_2 is the apt condition.
- By convention we make x bold since it is a vector.
- Thus our equation to predict the rent becomes: $y' = b + w_1x_1 + w_2x_2$

Summary of Linear Regression

- $y' = wx + b$
- Minimize mean squared error (or RMSE)
- In general have:
  $y' = b + w_1x_1 + \ldots + w_nx_n$
- Learn model weights $b, w_1, w_2, \ldots w_n$ from data