Calculus in AI and Machine Learning 1

Qiyam Tung

G-TEAMS
Department of Computer Science
Univeristy of Arizona

November 10, 2011

Introduction

Linear Regression is one of the tools we use in machine learning.

- Predicting house prices as a function of its size.
- Predicting consumption spending (a large number of input variables).
- Predicting the effect of factories on environments.

For each problem, find the parameters for the line given a set of points. If the line doesn't exist, explain why not.

1. (1,1) (2,2)
2. (2,2) (4, 10)
3. (1,7)
4. (1,1) (2,3) (3,2) (4,4)

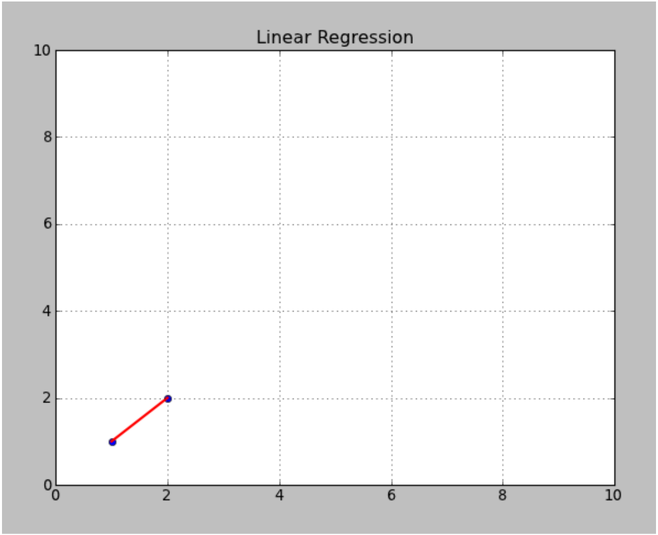


Figure: (1,1) (2,2)

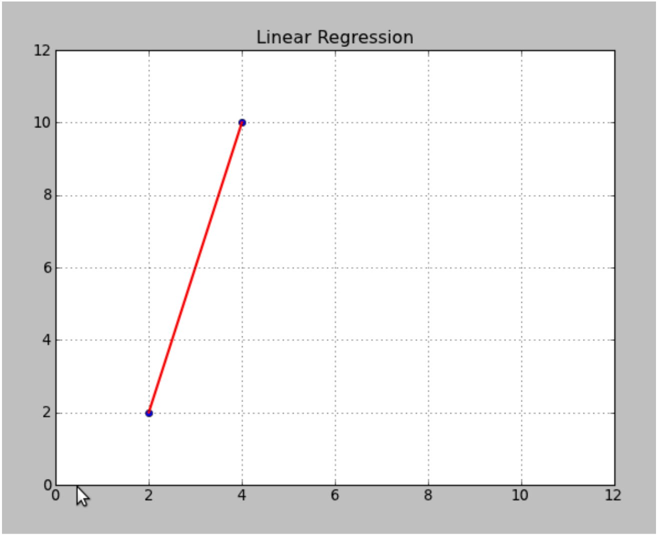


Figure: (2,2) (4,10)

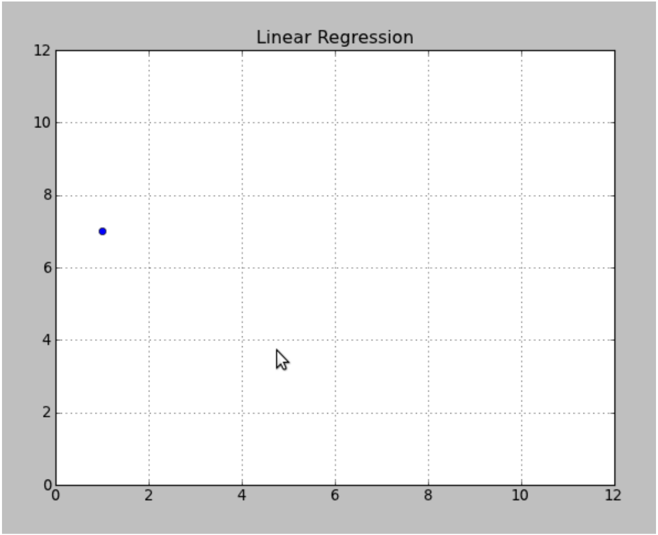


Figure: (1.7)

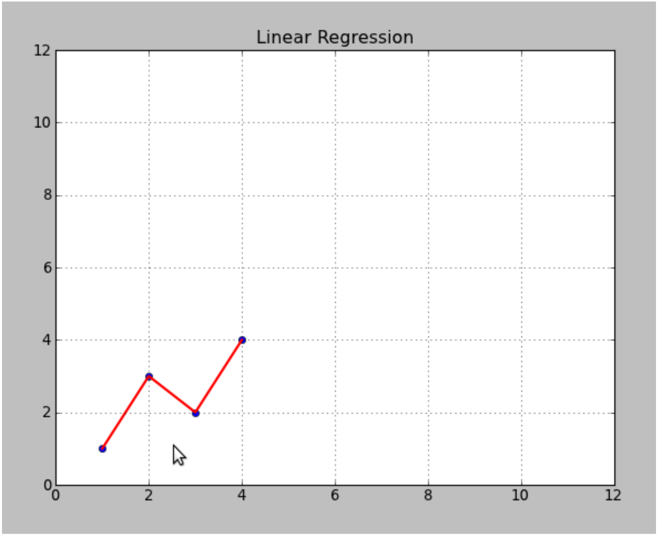


Figure: (1,1) (2,3) (3,2) (4,4)

How do you estimate a line?

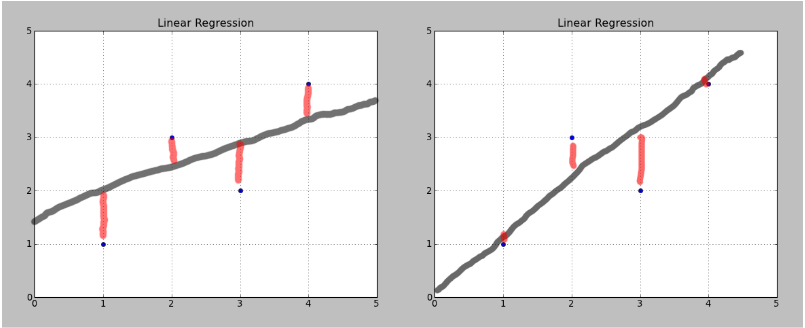


Figure: When we fit a line, we want to minimize the average distance of the points to the line we're fitting. The distance is indicated by the vertical lines from the point to the line.

$$Error = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{1}$$

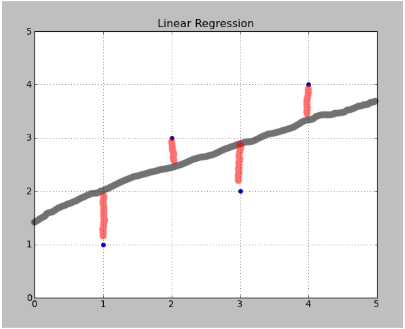
$$Error = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{1}$$

Scary!

$$Error = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \quad (1)$$

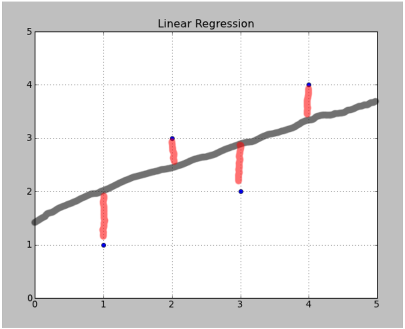
Scary!

Hold on...



$$h_\theta(x) - y = (\theta_0 + \theta_1 x) - y \tag{2}$$

Distance from line?



$$(h_\theta(x) - y)^2 = ((\theta_0 + \theta_1 x) - y)^2 \tag{3}$$

Squaring it always gives us a positive error. We don't want the errors to cancel out.

We want to minimize the **average distance** of the points to the **line** we're fitting. The **distance/error** is indicated by the vertical lines from the point to the line.

$$Error = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{4}$$

We want to minimize the **average distance** of the points to the **line** we're fitting. The **distance/error** is indicated by the vertical lines from the point to the line.

$$Error = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{4}$$

- Average: $\frac{1}{m}$

We want to minimize the **average distance** of the points to the **line** we're fitting. The **distance/error** is indicated by the vertical lines from the point to the line.

$$Error = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{4}$$

- Average: $\frac{1}{m}$
- "Distance" or mean-squared-error: $(h_\theta(x^{(i)}) - y^{(i)})^2$

We want to minimize the **average distance** of the points to the **line** we're fitting. The **distance/error** is indicated by the vertical lines from the point to the line.

$$Error = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{4}$$

- Average: $\frac{1}{m}$
- "Distance" or mean-squared-error: $(h_\theta(x^{(i)}) - y^{(i)})^2$
- Input: $x^{(1)}$, $x^{(2)}$, ..., $x^{(m)}$

We want to minimize the **average distance** of the points to the **line** we're fitting. The **distance/error** is indicated by the vertical lines from the point to the line.

$$Error = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \quad (4)$$

- Average: $\frac{1}{m}$
- "Distance" or mean-squared-error: $(h_\theta(x^{(i)}) - y^{(i)})^2$
- Input: $x^{(1)}$, $x^{(2)}$, ..., $x^{(m)}$
- Line: $h_\theta(x^{(i)})$

We want to minimize the **average distance** of the points to the **line** we're fitting. The **distance/error** is indicated by the vertical lines from the point to the line.

$$Error = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{4}$$

- Average: $\frac{1}{m}$
- "Distance" or mean-squared-error: $(h_\theta(x^{(i)}) - y^{(i)})^2$
- Input: $x^{(1)}$, $x^{(2)}$, ..., $x^{(m)}$
- Line: $h_\theta(x^{(i)})$
- Output: $y^{(1)}$, $y^{(2)}$, ..., $y^{(m)}$

$$Error = \frac{1}{4}\sum_{i=1}^{4}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{5}$$

$$Error = \frac{1}{4}\sum_{i=1}^{4}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{5}$$

$$\begin{aligned} Error \quad = \quad & \frac{1}{4}\Big((\theta_0 + \theta_1 x^{(1)} - y^{(1)})^2 + \\ & (\theta_0 + \theta_1 x^{(2)} - y^{(2)})^2 + \\ & (\theta_0 + \theta_1 x^{(3)} - y^{(3)})^2 + \\ & (\theta_0 + \theta_1 x^{(4)} - y^{(4)})^2\Big) \end{aligned}$$

$$Error = \frac{1}{4}\sum_{i=1}^{4}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{5}$$

$$\begin{aligned} Error \quad = \quad & \frac{1}{4}\Big((\theta_0 + \theta_1 x^{(1)} - y^{(1)})^2 + \\ & (\theta_0 + \theta_1 x^{(2)} - y^{(2)})^2 + \\ & (\theta_0 + \theta_1 x^{(3)} - y^{(3)})^2 + \\ & (\theta_0 + \theta_1 x^{(4)} - y^{(4)})^2\Big) \end{aligned}$$

$$Error = \frac{1}{4}((\theta_0+\theta_1-1)^2+(\theta_0+2\theta_1-3)^2+(\theta_0+3\theta_1-2)^2+(\theta_0+4\theta_1-4)^2) \tag{6}$$

Given the data $(x^{(1)}, y^{(1)})$, $(x^{(2)}, y^{(2)})$, $(x^{(3)}, y^{(3)})$, $(x^{(4)}, y^{(4)})$ or (1,1), (2,3), (3,2), (4,4), find the error of each of the lines with the given parameters.

1. $\theta_0 = 2$, $\theta_1 = 0$
2. $\theta_0 = 0$, $\theta_1 = 1$

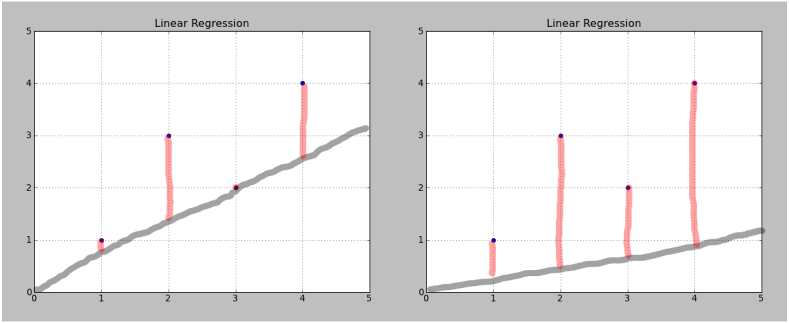


Figure: As we decrease the slope, the error increases dramatically

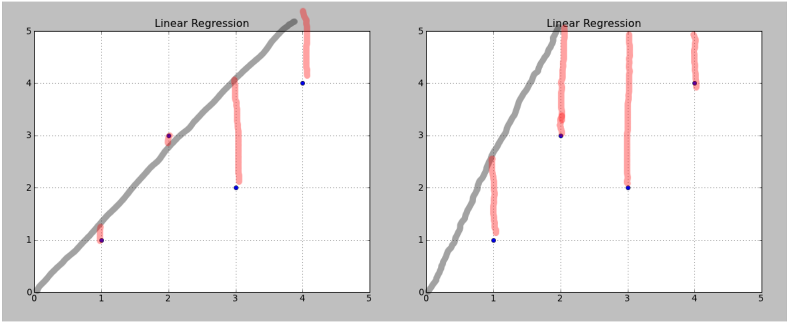


Figure: Similarly, if we increase the slope the error also increases dramatically

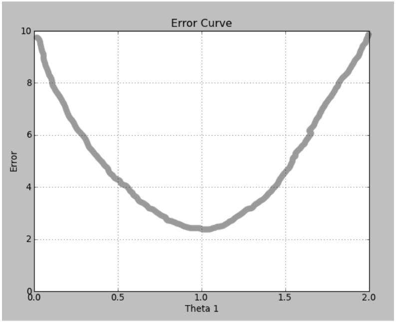


Figure: The error plotted as a function of theta.

How do you minimize the function? In other words, how do you find the minimum y for this error function?