CS 4104
APPLIED MACHINE LEARNING

Dr. Hashim Yasin

National University of Computer and Emerging Sciences,

Faisalabad, Pakistan.

LINEAR REGRESSION

Cost Function

Hypothesis:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Parameters:

$$\theta_0, \theta_1$$

Cost Function:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

Simplified:

$$h_\theta(x) = \theta_1 x$$

$$\theta_1$$

$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

$$\underset{\theta_1}{\text{minimize}}\ J(\theta_1)$$

What's next?

Have some function $J(\theta_0, \theta_1)$

Want $\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

Outline:

- Start with some random values θ_0, θ_1
- **Keep changing** θ_0, θ_1 to reduce $J(\theta_0, \theta_1)$

  until we hopefully end up at a minimum

Gradient Descent

Gradient descent algorithm

repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \quad \text{(simultaneously update } j = 0 \text{ and } j = 1\text{)}$$

}

Notice : α is the learning rate.

$$J(\theta_0,\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

Gradient Descent

Gradient descent algorithm

repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \quad (\text{for } j = 0 \text{ and } j = 1)$$

}

Correct: Simultaneous update

$$\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$
$$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$
$$\theta_0 := \text{temp0}$$
$$\theta_1 := \text{temp1}$$

Incorrect:

$$\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$
$$\theta_0 := \text{temp0}$$
$$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$
$$\theta_1 := \text{temp1}$$

Gradient Descent

Gradient descent algorithm

$$
\begin{aligned}
&\text{repeat until convergence } \{ \\
&\quad \theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad \text{(simultaneously update } j = 0 \text{ and } j = 1) \\
&\}
\end{aligned}
$$

?

Notice : α is the learning rate.

Gradient Descent

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Partial Derivative w.r.t. θ_0

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

$$\frac{\partial J(\theta_0, \theta_1)}{\partial \theta_0} = \frac{\partial}{\partial \theta_0} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

$$\frac{\partial J(\theta_0, \theta_1)}{\partial \theta_0} = \frac{1}{2m} \sum_{i=1}^{m} \frac{\partial}{\partial \theta_0} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)} \right)^2$$

$$\frac{\partial J(\theta_0, \theta_1)}{\partial \theta_0} = 2 \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)} \right) \frac{\partial}{\partial \theta_0} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)} \right)$$

$$\frac{\partial J(\theta_0, \theta_1)}{\partial \theta_0} = \frac{1}{m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)} \right) (1) = \boldsymbol{\frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)}$$

Gradient Descent

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Partial Derivative w.r.t. θ_1

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$\frac{\partial\, J(\theta_0, \theta_1)}{\partial \theta_1} = \frac{\partial}{\partial \theta_1} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$\frac{\partial\, J(\theta_0, \theta_1)}{\partial \theta_1} = \frac{1}{2m} \sum_{i=1}^{m} \frac{\partial}{\partial \theta_1} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^2$$

$$\frac{\partial\, J(\theta_0, \theta_1)}{\partial \theta_1} = 2\frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right) \frac{\partial}{\partial \theta_1} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)$$

$$\frac{\partial\, J(\theta_0, \theta_1)}{\partial \theta_1} = \frac{1}{m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right) x^{(i)} = \boldsymbol{\frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x^{(i)}}$$

Summary

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

Objective: $\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

Update rules:
$$\theta_0 := \theta_0 - \alpha\frac{\partial}{\partial\theta_0}J(\theta_0, \theta_1)$$
$$\theta_1 := \theta_1 - \alpha\frac{\partial}{\partial\theta_1}J(\theta_0, \theta_1)$$

Derivatives:

$$\frac{\partial}{\partial\theta_0}J(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)$$

$$\frac{\partial}{\partial\theta_1}J(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x^{(i)}$$

LINEAR REGRESSION WITH MULTIPLE VARIABLES

Multivariate Regression

Hypothesis: $h_\theta(x) = \theta^T x = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$

Parameters: $\theta_0, \theta_1, \ldots, \theta_n$

$x_0 = 1$

Cost function:

$$J(\theta_0, \theta_1, \ldots, \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

Gradient descent:

Repeat $\{$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \ldots, \theta_n)$$

$\}$ (simultaneously update for every $j = 0, \ldots, n$)

Gradient Descent

$$h_\theta\left(x^{(i)}\right) = \theta_0 + \theta_1 x_1{}^{(i)} + \theta_2 x_2{}^{(i)}$$

Partial Derivative w.r.t. θ_0

$$J(\theta_0, \theta_1, \theta_2) = \frac{\partial}{\partial \theta_0} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$\frac{\partial\, J(\theta_0, \theta_1, \theta_2)}{\partial \theta_0} = \frac{\partial}{\partial \theta_0} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$\frac{\partial\, J(\theta_0, \theta_1, \theta_2)}{\partial \theta_0} = \frac{1}{2m} \sum_{i=1}^{m} \frac{\partial}{\partial \theta_0} \left(\theta_0 + \theta_1 x_1{}^{(i)} + \theta_2 x_2{}^{(i)} - y^{(i)}\right)^2$$

$$\frac{\partial\, J(\theta_0, \theta_1, \theta_2)}{\partial \theta_0} = 2\frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x_1{}^{(i)} + \theta_2 x_2{}^{(i)} - y^{(i)}\right) \frac{\partial}{\partial \theta_0} \left(\theta_0 + \theta_1 x_1{}^{(i)} + \theta_2 x_2{}^{(i)} - y^{(i)}\right)$$

$$\frac{\partial\, J(\theta_0, \theta_1, \theta_2)}{\partial \theta_0} = \frac{1}{m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x_1{}^{(i)} + \theta_2 x_2{}^{(i)} - y^{(i)}\right) (1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)$$

Gradient Descent

$$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x_1{}^{(i)} + \theta_2 x_2{}^{(i)}$$

Partial Derivative w.r.t. θ_1

$$J(\theta_0, \theta_1, \theta_2) = \frac{\partial}{\partial \theta_0} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$\frac{\partial\, J(\theta_0, \theta_1, \theta_2)}{\partial \theta_1} = \frac{\partial}{\partial \theta_0} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$\frac{\partial\, J(\theta_0, \theta_1, \theta_2)}{\partial \theta_1} = \frac{1}{2m} \sum_{i=1}^{m} \frac{\partial}{\partial \theta_0} \left(\theta_0 + \theta_1 x_1{}^{(i)} + \theta_2 x_2{}^{(i)} - y^{(i)}\right)^2$$

$$\frac{\partial\, J(\theta_0, \theta_1, \theta_2)}{\partial \theta_1} = 2 \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x_1{}^{(i)} + \theta_2 x_2{}^{(i)} - y^{(i)}\right) \frac{\partial}{\partial \theta_1} \left(\theta_0 + \theta_1 x_1{}^{(i)} + \theta_2 x_2{}^{(i)} - y^{(i)}\right)$$

$$\frac{\partial\, J(\theta_0, \theta_1, \theta_2)}{\partial \theta_1} = \frac{1}{m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x_1{}^{(i)} + \theta_2 x_2{}^{(i)} - y^{(i)}\right) x_1^{(i)} = \boldsymbol{\frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x_1^{(i)}}$$

Gradient Descent

$$h_\theta(x^{(i)}) = \theta_0 + \theta_1 {x_1}^{(i)} + \theta_2 {x_2}^{(i)}$$

Partial Derivative w.r.t. θ_2

$$J(\theta_0, \theta_1, \theta_2) = \frac{\partial}{\partial \theta_0} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$\frac{\partial\, J(\theta_0, \theta_1, \theta_2)}{\partial \theta_2} = \frac{\partial}{\partial \theta_0} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$\frac{\partial\, J(\theta_0, \theta_1, \theta_2)}{\partial \theta_2} = \frac{1}{2m} \sum_{i=1}^{m} \frac{\partial}{\partial \theta_0} \left(\theta_0 + \theta_1 {x_1}^{(i)} + \theta_2 {x_2}^{(i)} - y^{(i)}\right)^2$$

$$\frac{\partial\, J(\theta_0, \theta_1, \theta_2)}{\partial \theta_2} = 2\frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 {x_1}^{(i)} + \theta_2 {x_2}^{(i)} - y^{(i)}\right) \frac{\partial}{\partial \theta_2} \left(\theta_0 + \theta_1 {x_1}^{(i)} + \theta_2 {x_2}^{(i)} - y^{(i)}\right)$$

$$\frac{\partial\, J(\theta_0, \theta_1, \theta_2)}{\partial \theta_2} = \frac{1}{m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 {x_1}^{(i)} + \theta_2 {x_2}^{(i)} - y^{(i)}\right) x_2^{(i)} = \boldsymbol{\frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x_2^{(i)}}$$

Gradient Descent

Previously $(n = 1)$:

Repeat $\{$

$$\theta_0 := \theta_0 - \alpha \underbrace{\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})}_{\frac{\partial}{\partial \theta_0} J(\theta)}$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x^{(i)}$$

(simultaneously update θ_0, θ_1)

$\}$

New algorithm $(n \geq 1)$:

Repeat $\{$

$$\theta_j := \theta_j - \alpha \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$$

(simultaneously update θ_j for $j = 0, \ldots, n$)

$\}$

$$\theta_0 := \theta_0 - \alpha \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_0^{(i)}$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_1^{(i)}$$

$$\theta_2 := \theta_2 - \alpha \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_2^{(i)}$$

$\ldots$

LINEAR REGRESSION WITH MULTIPLE VARIABLES(MATRIX FORM)

EXAMPLES

Example 1

Examples: $m = 4.$

x_0	Size (feet²) x_1	Number of bedrooms x_2	Number of floors x_3	Age of home (years) x_4	Price ($1000) y
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1534	3	2	30	315
1	852	2	1	36	178

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{bmatrix} \qquad y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

$$\theta = (X^T X)^{-1} X^T y$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)^T$

simultaneously update

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

Example 2

A chemical process expects the yield to be affected by two factors x_1 and x_2

Observations recorded for these two factors are shown in the given table.

Observation Number	Factor 1 (x_{i1})	Factor 2 (x_{i2})	Yield (y_i)
1	41.9	29.1	251.3
2	43.4	29.3	251.3
3	43.9	29.5	248.3
4	44.5	29.7	267.5
5	47.3	29.9	273.0
6	47.5	30.3	276.5
7	47.9	30.5	270.3
8	50.2	30.7	274.9
9	52.8	30.8	285.0
10	53.2	30.9	290.0
11	56.7	31.5	297.0
12	57.0	31.7	302.5
13	63.5	31.9	304.5
14	65.3	32.0	309.3
15	71.1	32.1	321.7
16	77.0	32.5	330.7
17	77.8	32.9	349.0

Example 2

- The first order regression model is,

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$$\boldsymbol{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

$$X = \begin{bmatrix} 1 & 41.9 & 29.1 \\ 1 & 43.4 & 29.3 \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & 77.8 & 32.9 \end{bmatrix} \quad y = \begin{bmatrix} 251.3 \\ 251.3 \\ . \\ . \\ . \\ 349.0 \end{bmatrix}$$

Example 2

$$\boldsymbol{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 17 & 941 & 525.3 \\ 941 & 54270 & 29286 \\ 525.3 & 29286 & 16254 \end{bmatrix}^{-1} \begin{bmatrix} 4902.8 \\ 276610 \\ 152020 \end{bmatrix}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} -153.51 \\ 1.24 \\ 12.08 \end{bmatrix}$$

Example 2

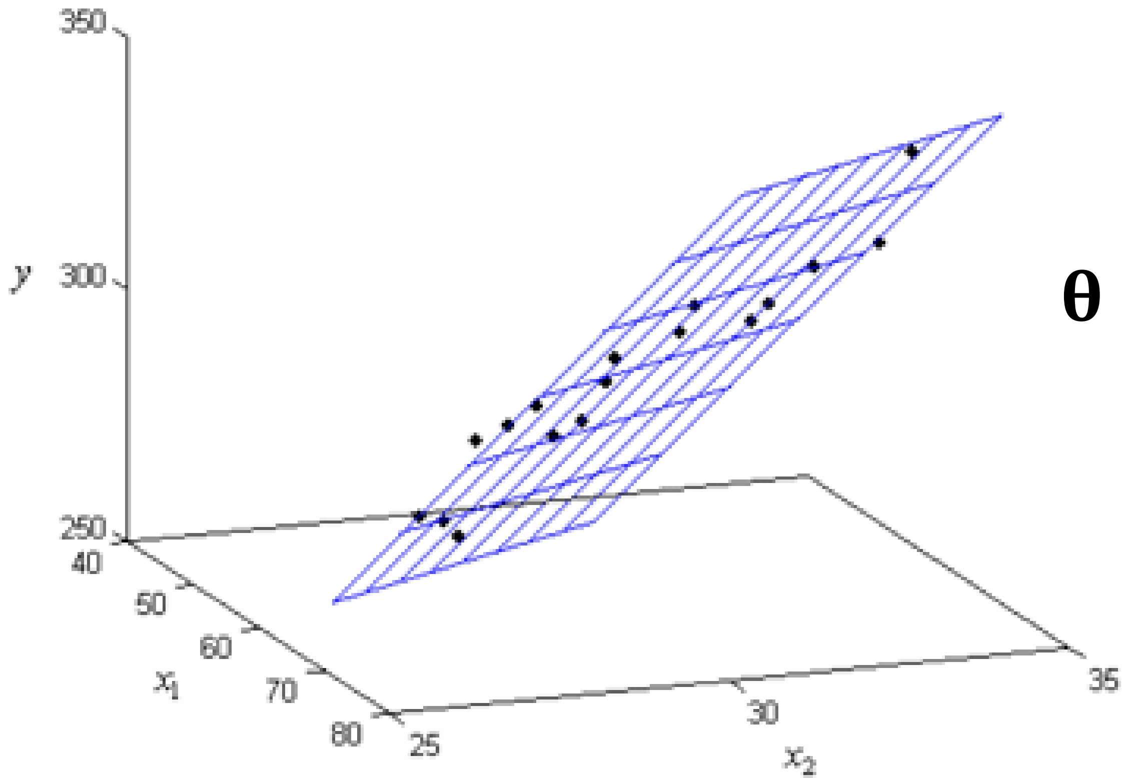
$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta_0} \\ \boldsymbol{\theta_1} \\ \boldsymbol{\theta_2} \end{bmatrix} = \begin{bmatrix} 17 & 941 & 525.3 \\ 941 & 54270 & 29286 \\ 525.3 & 29286 & 16254 \end{bmatrix}^{-1} \begin{bmatrix} 4902.8 \\ 276610 \\ 152020 \end{bmatrix}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta_0} \\ \boldsymbol{\theta_1} \\ \boldsymbol{\theta_2} \end{bmatrix} = \begin{bmatrix} -153.51 \\ 1.24 \\ 12.08 \end{bmatrix}$$

$$\boxed{h_\theta(x) = -153.51 + 1.24x_1 + 12.08x_2}$$

Example 2

$$h_\theta(x) = -153.51 + 1.24x_1 + 12.08x_2$$



$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta_0} \\ \boldsymbol{\theta_1} \\ \boldsymbol{\theta_2} \end{bmatrix} = \begin{bmatrix} -153.51 \\ 1.24 \\ 12.08 \end{bmatrix}$$

Acknowledgement

Tom Mitchel, Russel & Norvig, Andrew Ng, Alpydin & Ch. Eick.