KHOA CNTT & TRUYỀN THÔNG
BM KHOA HỌC MÁY TÍNH

HỒI QUY
REGRESSION

Giáo viên giảng dạy:
TS. TRẦN NGUYỄN MINH THƯ
tnmthu@cit.ctu.edu.vn

1

Quy ước

- Biến **đầu vào** (input variables)/đặc trưng (features), kí hiệu: $x^{(i)}$
- Biến **đầu ra** (output variable)/biến mục tiêu, kí hiệu y(i)
- Mẫu huấn luyện (training example)

  kí hiệu **($X^{(i)}$, $y^{(i)}$)**
- Tập huấn luyện **$X = \{(x^{(i)}, y^{(i)})\}$**, i = 1..m

Square meters	Bedrooms	Floors	Age of building (years)	Price in 1000€
x1	*x2*	*x3*	*x4*	*y*
200	5	1	45	460
131	3	2	40	232
142	3	2	30	315
756	2	1	36	178
…	…	…	…	…

e

$$x^{(3)} = \begin{bmatrix} 142 \\ 3 \\ 2 \\ 30 \end{bmatrix}$$

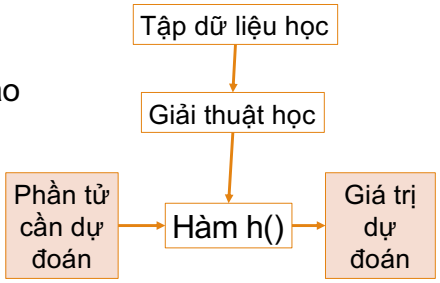
$$x_1^{(4)} = 756$$

2

Phân loại học máy – **học có giám sát**

Từ tập dữ liệu huấn luyện $\{(x_1, y_1), (x_2,y_2),..,(x_m,y_m)$

- Tìm hàm h (hypothesis) X=>Y sao cho h(x) dự báo được **y** từ **x**
- **Y là giá trị liên tục:** sử dụng pp hồi quy (regression)
- **Y là giá trị rời rạc:** sử dụng pp phân lớp (classification)



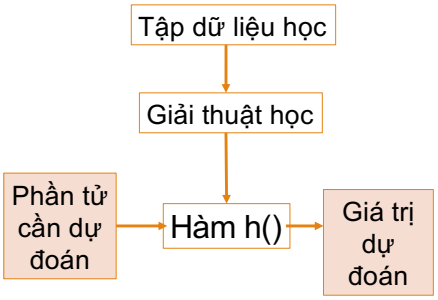
3

3

Phân loại học máy – **học có giám sát**

Ví dụ: bài toán dự báo giá nhà

Living area (feet^2)	Price (1000$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮



– **Đầu vào/thuộc tính**: ?
– **Đầu ra**: ?

Xác định :
Dự báo cái gì?
Dựa trên thông tin gì?
Giải thuật gì?

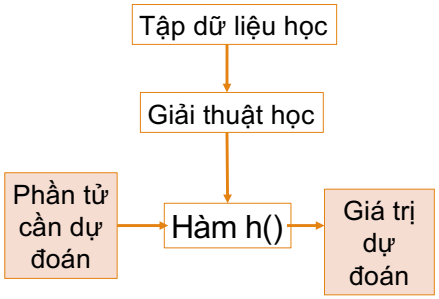
4

4

Phân loại học máy – **học có giám sát**

Ví dụ: bài toán dự báo giá nhà

Living area ($feet^2$)	Price (1000$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮



– **Đầu vào/thuộc tính**: diện tích
– **Đầu ra**: giá nhà - **giá trị liên tục**

Xác định thuộc tính:
Dự báo cái gì? Y = giá nhà
Dựa trên thông tin gì?
Giải thuật gì? **Hồi quy**

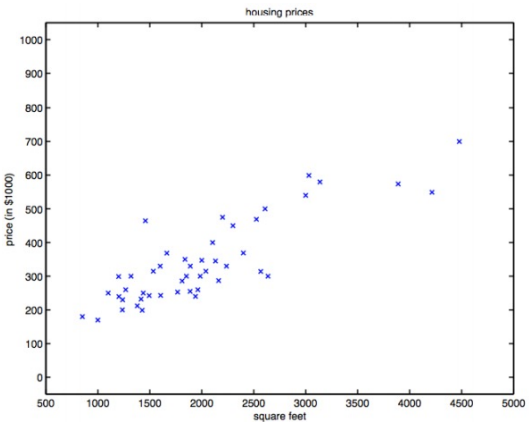
5

5

Ví dụ dự đoán giá nhà

Dự báo giá nhà dựa vào diện tích

Living area ($feet^2$)	Price (1000$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮

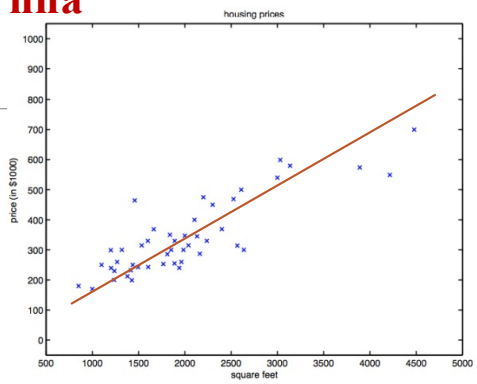


6

Ví dụ dự đoán giá nhà

Dự báo giá nhà dựa vào diện tích

Living area ($feet^2$)	Price (1000$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮



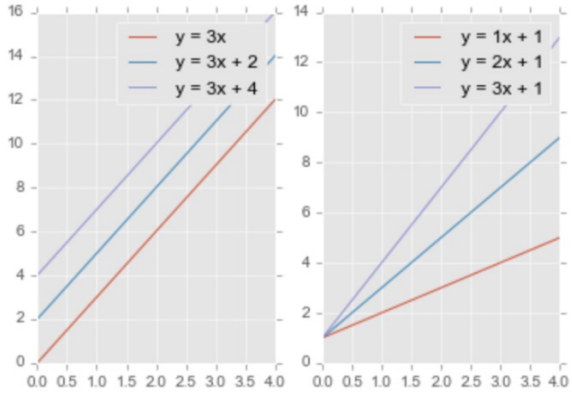
➢**Biểu diễn giả thiết (hàm dự báo) h**

– Ví dụ h là một hàm tuyến tính 1 biến, h(x1) có dạng: $h_\theta(x) = \theta_0 + \theta_1 x_1$

Trong đó, θ_0, θ_1 là các tham số cần mà ta phải tìm trong quá trình "dạy cho máy học" hay còn gọi là quá trình huấn luyện.

7

Phương trình đường thẳng: y= ax + b



- θ_0 b quyết định điểm giao của đường thẳng với trục y, intercept/observation noise: điểm mà đường thẳng cắt trục Y.
- θ_1 a quyết định góc của đường thẳng – slope/coefficients: độ dốc của đường thẳng h(x)

8

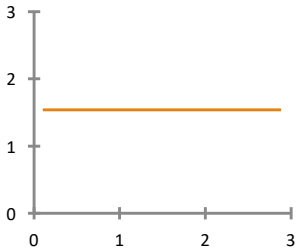
8

Phương trình đường thẳng

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

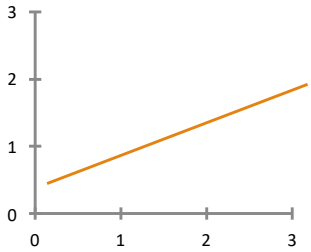
Cần xác định các tham số: θ_i

$h_\theta(x) = 1.5 + 0.x$



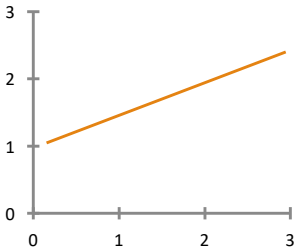
$\theta_0 = 1.5$
$\theta_1 = 0$

$h_\theta(x) = 0.5x$



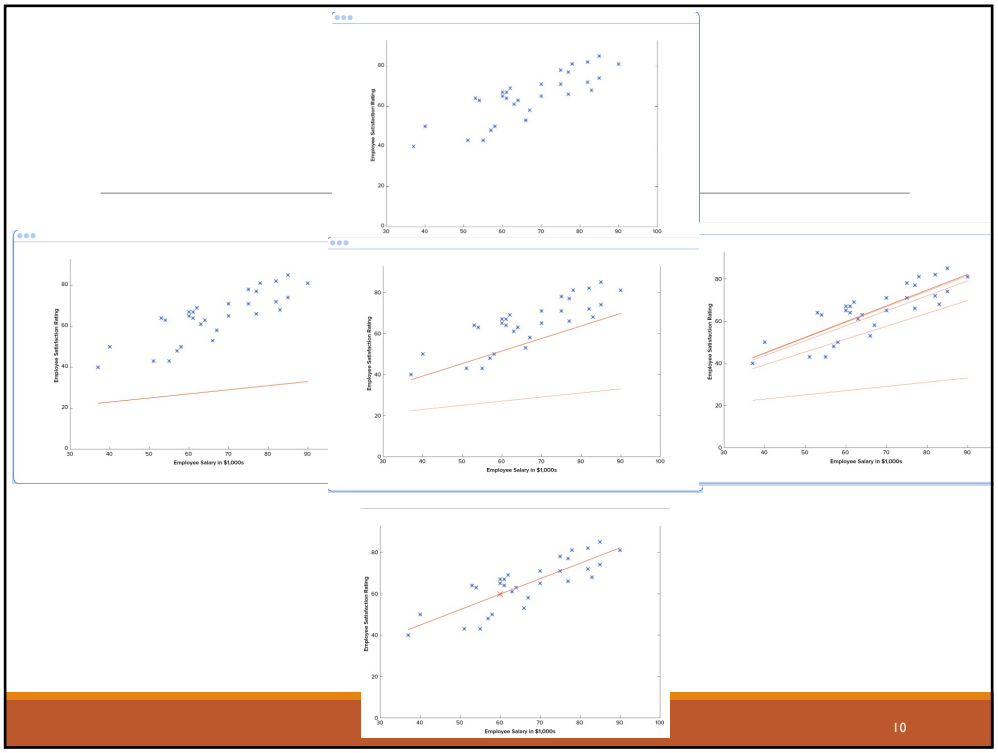
$\theta_0 = 0$
$\theta_1 = 0.5$

$h_\theta(x) = 1 + 0.5x$



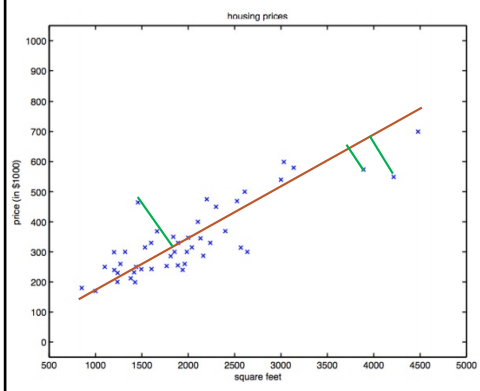
$\theta_0 = 1$
$\theta_1 = 0.5$

9



10

Ví dụ dự đoán giá nhà



Living area ($feet^2$)	Price (1000$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮

Trong khi sử dụng hồi quy tuyến tính, mục tiêu của chúng ta là để làm sao một đường thẳng có thể tạo được sự phân bố gần nhất với hầu hết các điểm. Do đó làm giảm khoảng cách (sai số) của các điểm dữ liệu cho đến đường đó.

11

Hồi quy tuyến tính

– Ta phải tìm hàm $h_\theta(x^{(i)})$ sao cho **h(x)** gần với **y** nhất (sai số dự đoán)

Nói cách khác, chúng ta muốn giá trị sau đây càng nhỏ càng tốt:

$$h_\theta(x^{(i)}) - y^{(i)}$$

12

Hồi quy tuyến tính

– Ta phải tìm hàm $h_\theta(x^{(i)})$ sao cho **h(x)** gần với **y** nhất (sai số dự đoán)

Nói cách khác, chúng ta muốn giá trị sau đây càng nhỏ càng tốt:

$$h_\theta(x^{(i)}) - y^{(i)}$$

– Hàm chi phí/hàm lỗi (cost function/error function) của **m** phần tử

- Hàm lỗi **sai số tuyệt đối**: $\sum_{i=1}^{m} \left| h_\theta(x^{(i)}) - y^{(i)} \right|$
- Hàm lỗi **sai số bình phương**: $\sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$

13

Hồi quy

Dạy cho máy học/huấn luyện như thế nào ?

– Tìm các tham số θ từ tập huấn luyện sao cho lỗi huấn luyện nhỏ nhất.
– Ta phải tìm h sao cho **h(x)** gần với **y** nhất = $h_\theta(x^{(i)}) - y^{(i)}$
– Hàm chi phí/hàm lỗi (cost function/error function)

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

Mục tiêu tìm θ sao cho J(θ) nhỏ nhất

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

14

Giảm gradient

- Mục tiêu: Tìm θ sao cho J(θ) nhỏ nhất
  - Khởi tạo ngẫu nhiên θ
  - Tăng/giảm θ một lượng Δθ sao cho J(θ +/- Δθ) nhỏ hơn J(θ)

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

α: tốc độ học

LMS (Least mean square): bình phương trung bình nhỏ nhất

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

15

Giải thuật LMS

Tính đạo hàm riêng theo từng tham số:

$$\begin{aligned}
\frac{\partial}{\partial \theta_j} J(\theta) &= \frac{\partial}{\partial \theta_j} \frac{1}{2} (h_\theta(x) - y)^2 \\
&= 2 \cdot \frac{1}{2} (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j} (h_\theta(x) - y) \\
&= (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j} \left(\sum_{i=0}^{n} \theta_i x_i - y \right) \\
&= (h_\theta(x) - y)\, x_j
\end{aligned}$$

Đạo hàm theo θj

16

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

Giải thuật LMS

Tính đạo hàm riêng theo từng tham số:

$$\begin{aligned}
\frac{\partial}{\partial \theta_j} J(\theta) &= \frac{\partial}{\partial \theta_j} \frac{1}{2} (h_\theta(x) - y)^2 \\
&= 2 \cdot \frac{1}{2} (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j} (h_\theta(x) - y) \\
&= (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j} \left(\sum_{i=0}^{n} \theta_i x_i - y \right) \\
\frac{\partial}{\partial \theta_j} J(\theta) &= (h_\theta(x) - y)\, x_j
\end{aligned}$$

17

Đạo hàm riêng

Scalar multiple rule: $\frac{d}{dx}(\alpha u) = \alpha \frac{du}{dx}$

Sum rule: $\frac{d}{dx} \sum u = \sum \frac{du}{dx}$

Power rule: $\frac{d}{dx} u^n = n u^{n-1} \frac{du}{dx}$

Chain rule: $\frac{d}{dx} f(g(x)) = f'\big(g(x)\big) g'(x)$

$$\frac{d}{d\theta_1} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right) = \frac{d}{d\theta_1} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)} \right) = x^{(i)}$$

18

18

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

Giải thuật LMS

$$\frac{\partial}{\partial \theta_j} J(\theta) = (h_\theta(x) - y)\, x_j$$

Nếu chỉ có **1 mẫu huấn luyện**, ta cập nhật:

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

Nếu có **nhiều mẫu huấn luyện**, sử dụng luật cập nhật:

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

Hoặc:

for i=1 to m, {

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

}

19

$$\frac{\partial}{\partial \theta_j} J(\theta) = (h_\theta(x) - y)\, x_j$$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

Giải thuật LMS

Sử dụng luật cập nhật theo Batch Gradient descent (GD)

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

Hoặc Stochastic gradient descent (SGD)

for i=1 to m, {

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

}

20

Ví dụ

Cho tập dữ liệu gồm 3 phần tử như bảng bên, hãy thực hiện các công việc sau

- Tìm hàm hồi quy h(x) với giá trị khởi tạo (0, 1), tốc độ học: 0.2, số bước lặp: 2
- Dự đoán giá trị y cho phần tử có **x = 3**

x	y
1	2
2	3
4	6

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

21

21

Ví dụ

Cho tập dữ liệu gồm 3 phần tử như bảng bên, hãy thực hiện các công việc sau

- Biểu diễn tập dữ liệu lên mặt phẳng toạ độ Oxy
- Tìm hàm hồi quy h(x) với giá trị khởi tạo (0, 1), tốc độ học: 0.2, số bước lặp: 2
- Vẽ đường hồi quy lên mặt phẳng toạ độ
- Dự đoán giá trị y cho phần tử có **x = 3**

x	y
1	2
2	3
4	6

for i=1 to m, {

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

}

22

22

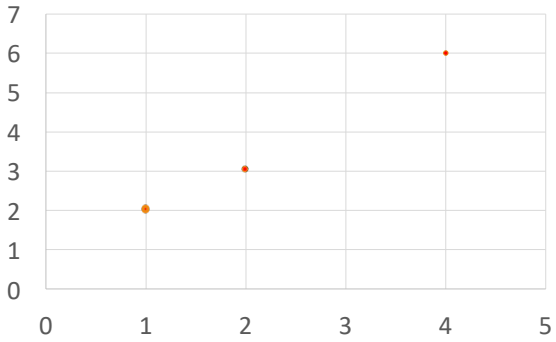
Ví dụ

for i=1 to m, {

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

}

- Biểu diễn tập dữ liệu lên mặt phẳng toạ độ Oxy



x	y
1	2
2	3
4	6

23

23

Ví dụ

for i=1 to m, {

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

}

- Tìm hàm hồi quy h(x) với giá trị khởi tạo (0, 1), tốc độ học: 0.2, số bước lặp: 2

m=?

j=?

Có bao nhiêu giá trị θ

$h_\theta(x)$ = ?

x	y
1	2
2	3
4	6

24

24

Ví dụ

for i=1 to m, {

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

}

- Tìm hàm hồi quy h(x) với giá trị khởi tạo (0, 1), tốc độ học: 0.2, số bước lặp: 2

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

$\theta_0 = 0, \theta_1 = 1; \alpha = 0.2 \quad h_\theta(x)_0 = 0 + 1.x$

Lần lặp 1:

Phần tử thứ 1(x=1,y=2):($x^{(1)}$):

x	y
1	2
2	3
4	6

Tìm θ_0 $\theta_0 = \theta_0 + \alpha*(y^1 - \{0 + 1* x^1_1\})*x^1_0$

$= 0 + 0.2(2- \{0 +1*1\})*1 = 0.2$

Tìm θ_1 $\theta_1 = \theta_1 + \alpha*(y^1 - \{0 + 1* x^1_1\})* x^1_1$

$= 1 + 0.2(2- \{0 +1*1\})*1 = 1.2$

25

25

Ví dụ

for i=1 to m, {

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

}

- Tìm hàm hồi quy h(x) với giá trị khởi tạo (0, 1), tốc độ học: 0.2, số bước lặp: 2

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

$\theta_0 = 0, \theta_1 = 1; \alpha = 0.2 \quad h_\theta(x)_0 = 0 + 1.x$

Lần lặp 1:

Pt 2(2,3): ($x^{(2)}$):

x	y
1	2
2	3
4	6

Tìm θ_0: $\theta_0 = \theta_0 + \alpha*(y^2 - \{\theta_0 + \theta_1 *x^{(2)}_1\})*x^{(2)}_0$

$= \mathbf{0.2} + 0.2(3 - \{0.2+\mathbf{1.2}*2\}) *1 = \mathbf{0.28}$

Tìm θ_1 $\theta_1 = \theta_1 + \alpha*(y^2 - \{\theta_0 + \theta_1 *x^{(2)}_1\})*x^{(2)}_1$

$= 1.2 + 0.2*(3 - (0.2+1.2*2)) *2 = \mathbf{1.36}$

26

26

Ví dụ

for i=1 to m, {

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

}

- Tìm hàm hồi quy h(x) với giá trị khởi tạo (0, 1), tốc độ học: 0.2, số bước lặp: 2

$h_\theta(x) = \theta_0 + \theta_1 x$; $\theta_0 = 0$, $\theta_1 = 1$; $\alpha = 0.2$ $h_\theta(x)_0 = 0 + 1.x$

x	y
1	2
2	3
4	6

Lần lặp 1: Tìm θ_0

Pt 3(4,6): $(x^{(3)})$

Tìm θ_0: $\theta_0 = \theta_0 + \alpha*(y^3 - (\theta_0 + \theta_1 * x^3{}_1))* x^3{}_0$

$= 0.28 + 0.2*(6 - (0.28 + 1.36*4))*1 = 0.336$

Tìm θ_1 $\theta_1 = \theta_1 + \alpha*(y^3 - \{\theta_0 + \theta_1 * x^3{}_1\})* x^3{}_1$

$= 1.36 + 0.2(6 - (0.28 + 1.36*4))*4 = 1.58$

$h_\theta(x)_0 = 0.336 + 1.58.x$

27

27

Ví dụ

for i=1 to m, {

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

}

- Tìm hàm hồi quy h(x) với giá trị khởi tạo (0, 1), tốc độ học: 0.2, số bước lặp: 2

$h_\theta(x) = \theta_0 + \theta_1 x \Rightarrow h_\theta(x)_0 = 0.336 + 1.58.x$

x	y
1	2
2	3
4	6

Lần lặp 2: Tìm θ_0

Tiếp tục với giá trị $\theta_0 = 0.336$, $\theta_1 = 1.58$;

28

28