

# Linear Regression



VietAI Teaching Team

# Nội dung

1. Kí hiệu (Notation)
2. Bài toán
3. Linear Regression Model
4. Cost function
5. Gradient Descent
6. Mở rộng
7. Normal Equation (đọc thêm)

# 1 Kí hiệu

- $m$  : số lượng mẫu dữ liệu trong tập huấn luyện  $\{(x^{(i)}, y^{(i)}) ; i = 1, \dots, m\}$
- $n$  : số chiều của input vector
- $x^{(i)}$  : input vector (**feature vector**)  $x^{(i)} \in \mathbb{R}^n$
- $X$  : ma trận với mỗi hàng là input vector  $x^{(i)T}$
- $y$  : vector chứa các giá trị output tương ứng với các input vectors
- $w$  : vector tham số/trọng số (**parameter/weight**)  $w \in \mathbb{R}^n$
- $J$  : cost function
- $\nabla J$  : vector đạo hàm (theo  $w$ ) của  $J$
- $\alpha$  : learning rate (trong thuật toán Gradient Descent)

## 2 Bài toán

Area - $x_1$	No of bedrooms - $x_2$	Price - $y$
2104	3	400
1600	3	330
2400	3	369
1416	2	232

Từ dữ liệu giá nhà đã thu thập được tại một thành phố, xây dựng mô hình dự đoán. Cho dữ liệu  $x = (x_1, x_2)$  mới, ước lượng giá trị  $y$  tương ứng.

### 3 Linear Regression Model

- Cho tập dữ liệu:  $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})$
- Tìm hàm  $y = f(x)$  phù hợp nhất mô hình hóa được tập dữ liệu trên.

$$h_w(x) = \sum_{i=1}^n w_i x_i + b = w^\top x + b$$

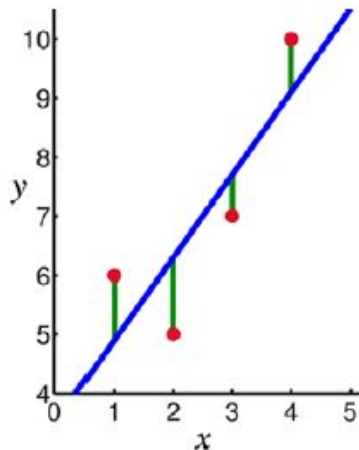
## 4 Cost function

- Cho tập dữ liệu:  $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})$

$$\hat{y} = h_w(x) = \sum_{i=1}^n w_i x_i + b = w^\top x + b$$

- Cost function được định nghĩa như sau:

$$\begin{aligned} J(w) &= \frac{1}{2m} \sum_{i=1}^m \text{loss}(\hat{y}^{(i)}, y^{(i)}) = \frac{1}{2m} \sum_{i=1}^m (\hat{y}^{(i)} - y^{(i)})^2 \\ &= \frac{1}{2m} \sum_{i=1}^m (w^\top x^{(i)} + b - y^{(i)})^2 \end{aligned}$$





## 5 Gradient Descent

- Lặp đến khi hội tụ công thức:

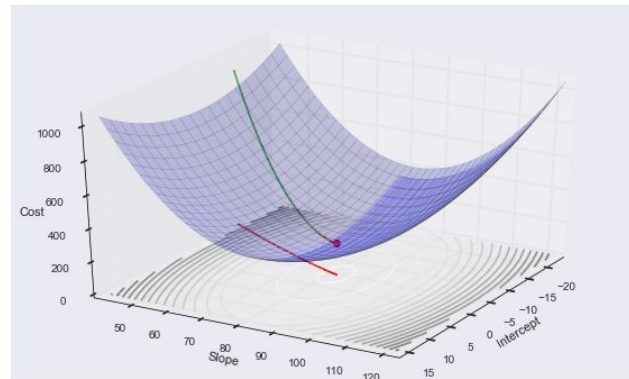
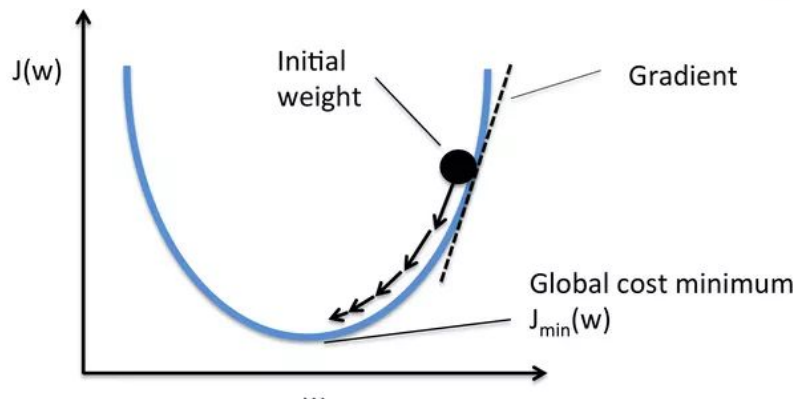
$$w_{new} = w_{old} - \alpha \nabla J(w_{old})$$

- Ví dụ:

$$f(w) = w^2 - 4w + 5$$

$$w_0 = 0$$

$$\alpha = 0.1$$



## 5 Gradient Descent

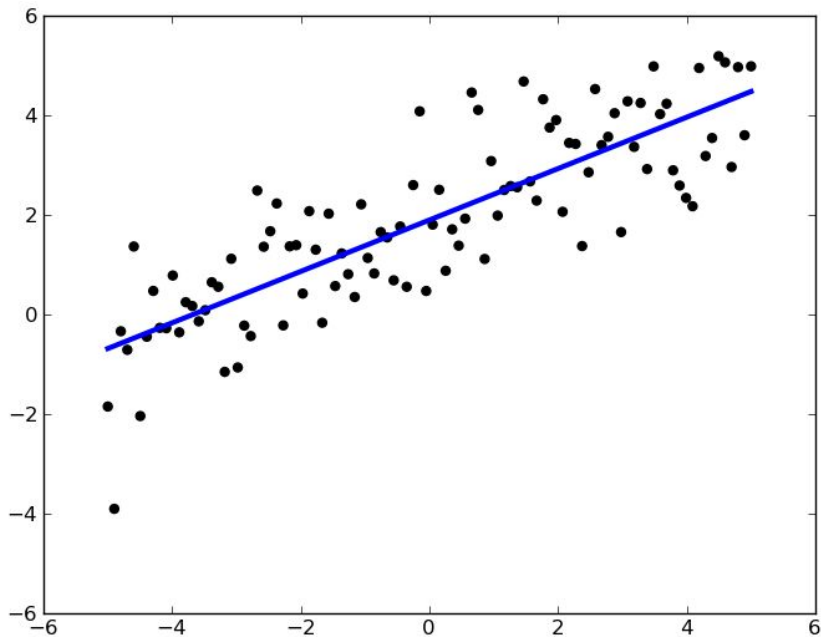
- Lặp đến khi hội tụ công thức:

$$w_{new} = w_{old} - \alpha \nabla J(w_{old})$$

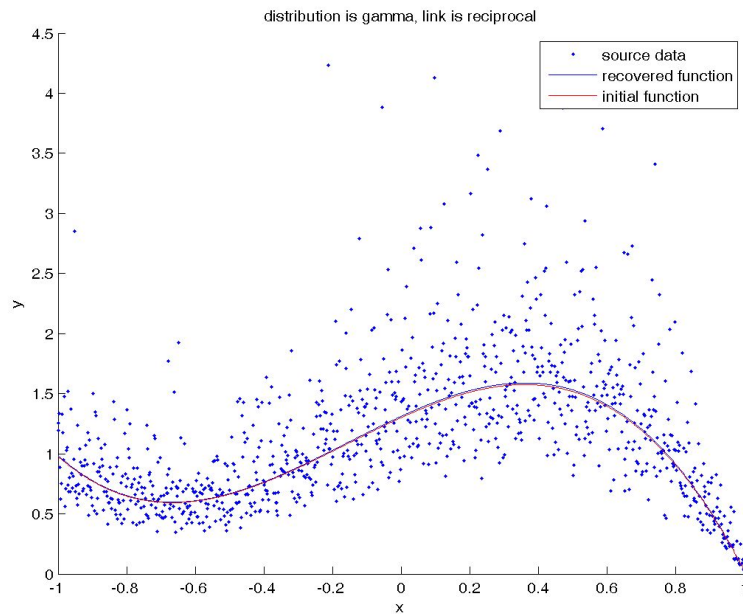
- Đạo hàm riêng của các phần tử trong  $w$

$$\frac{\partial J(w)}{\partial w_j} = \frac{1}{m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)}) x_j^{(i)}$$





Linear model



Linear model ???

## 7 Normal Equation (đọc thêm)

- Viết lại công thức cost function:

$$J(w) = \frac{1}{2m} \|Xw - y\|^2$$

- Viết lại công thức tính gradient:

$$\nabla J(w) = \frac{1}{m} X^\top (Xw - y)$$

- Giải phương trình gradient = 0

$$\frac{1}{m} X^\top (Xw - y) = 0$$

- Thu được kết quả

$$w = (X^\top X)^{-1} X^\top y$$

# Tài liệu tham khảo

1. Chapter 3 - Linear Regression - Pattern Recognition and Machine Learning  
C.M.Bishop
2. House Price data:  
[https://github.com/amaas/stanford\\_dl\\_ex/blob/master/ex1/housing.data](https://github.com/amaas/stanford_dl_ex/blob/master/ex1/housing.data)
3. Machine Learning: Bayesian and Optimization Perspective - Sergios Theodoridis
4. Linear Regression - Machine Learning Course on Coursera by Andrew Ng