Hồi Qui Tuyến Tính

TS. Nguyễn Thanh Hiên
Email: hien.nguyen@newai.vn

Trợ giảng 1: ThS. Dương Hữu Phúc
Email: phuc.duong@newai.vn

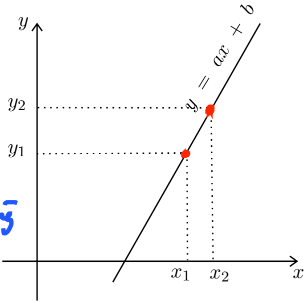
Trợ giảng 2: KS. Nguyễn Tân Kim
Email: kim.nguyen@newai.vn

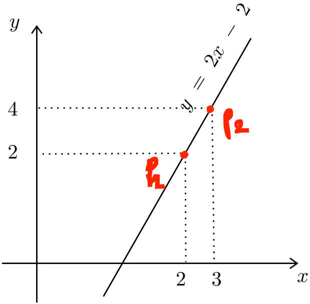
https://course.newai.vn

https://blog.newai.vn

Hàm tuyến tính

Xét trong không gian hai chiều hàm tuyến tính là đường thẳng dạng $y = ax + b$



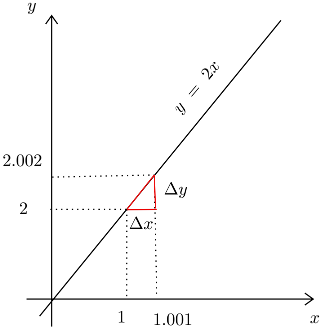


Cho hai điểm $P_1(2,2)$, $P_2(3,4)$ ta sẽ tìm được đường thẳng $y = ax + b$ với a, b là các tham số

$$\begin{cases} 2 = 2a + b \\ 4 = 3a + b \end{cases} \Rightarrow \begin{cases} a = 2 \\ b = -2 \end{cases}$$

Như vậy: P_1 và P_2 là dữ liệu, a và b là các tham số

Độ dốc



- Độ dốc (slope) hay còn gọi là gradient: $slope = \dfrac{\Delta y}{\Delta x} = \dfrac{y_2 - y_1}{x_2 - x_1}$

Siêu phẳng

- Khi $\boldsymbol{x}$ là véc-tơ n chiều, ta có siêu phẳng (hyperplane) $y = \boldsymbol{w}^T\boldsymbol{x} + b$ trong không gian $n+1$ chiều

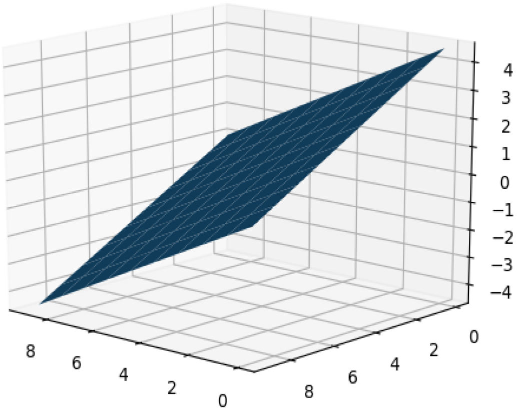
$$\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}, \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- Ta cũng có thể viết như sau

$$y = \begin{bmatrix} w_0 & w_1 & \dots & w_n \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$, với $b = w_0$ vì $x_0 = 1$

Siêu phẳng

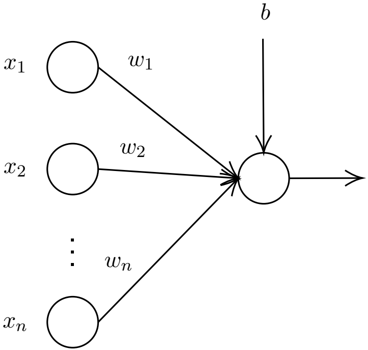
- Ví dụ siêu phẳng trong không gian 3 chiều



Hồi qui tuyến tính

- Tập huấn luyện có m mẫu: $\{(\boldsymbol{x}^{(i)}, y^{(i)}\}_{i=\overline{1,m}}$
  - Mỗi cặp $(\boldsymbol{x}, y)$ là một mẫu huấn luyện
  - $\boldsymbol{x}$: véc-tơ đặc trưng n chiều
  - y: kết quả
    - y rời rạc: bài toán phân lớp
    - y liên tục: bài toán hồi qui
- Hồi qui tuyến tính: $\hat{y} = h_{\boldsymbol{\theta}}(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b$, $\quad \boldsymbol{\theta} = \{\boldsymbol{w}, b\}$
- Hàm mất mát (loss function):
  - Ước lượng sự sai biệt giữa $\hat{y}$ và y
  - Kí hiệu: $\mathcal{L}(\hat{y}, y)$

Hồi qui tuyến tính



$$\begin{aligned}\hat{y} &= \boldsymbol{w}^T\boldsymbol{x} + b \\ &= \sum_{i=1}^{n} w_i x_i + b\end{aligned}$$

$$\hat{y} = h_{\boldsymbol{\theta}}(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b = \begin{bmatrix} w_1 & w_2 & \dots & w_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + b = \sum_{i=1}^{n} w_i x_i + b$$

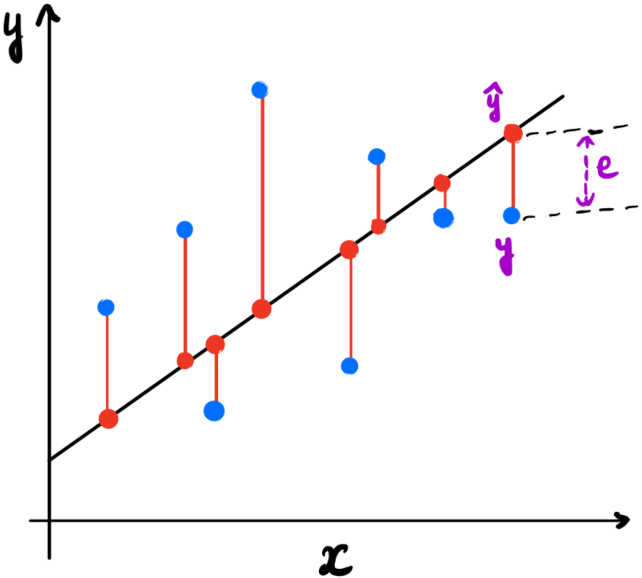
Hàm chi phí

- Tập huấn luyện: $\{(\boldsymbol{x}^{(i)}, y^{(i)})\}_{i=1}^{m}$
- Hàm chi phí (cost function):

$$J(\boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\mathcal{L}(\hat{y}^{(i)}, y^{(i)}) \qquad \hat{y}^{(i)} = f(\boldsymbol{x}^{(i)}; \boldsymbol{\theta}) \qquad \boldsymbol{\theta} = \{\boldsymbol{w}, b\}$$

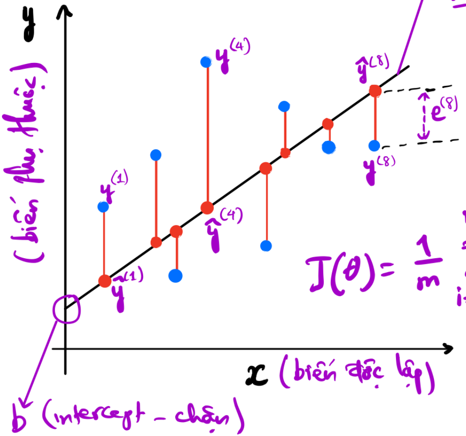
- Hàm mất mát: $\mathcal{L}(\hat{y}, y)$
  - Hàm bình phương sai biệt (square error): $\mathcal{L}(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$

Hồi qui tuyến tính



Hồi qui tuyến tính

$$\hat{y} = h_\theta(x) = w^T x + b, \quad \theta = \{w, b\}$$



$$e^{(8)} = \frac{1}{2}\left(\hat{y}^{(8)} - y^{(8)}\right)^2$$

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

Mục tiêu : tìm θ sao cho $J(\theta)$ cực tiểu

Maximum Likelihood

Tập huấn luyện: $\{(\boldsymbol{x}^{(i)}, y^{(i)})\}_{i=1}^{m}$

Giải sử các mẫu i.i.d: independent and identically distributed

$$y^{(i)} = \hat{y}^{(i)} + \epsilon^{(i)} = h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) + \epsilon^{(i)}$$

Giả sử: $\epsilon^{(i)} \sim \mathcal{N}(0, \sigma^2)$

$$\begin{aligned} p(\epsilon^{(i)}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\epsilon^{(i)}\right)^2}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y^{(i)} - \hat{y}^{(i)}\right)^2}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y^{(i)} - h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})\right)^2}{2\sigma^2}\right) \end{aligned}$$

$\Rightarrow p(y^{(i)} | \boldsymbol{x}^{(i)}; \boldsymbol{\theta}) \sim \mathcal{N}\left(h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}), \sigma^2\right)$

Maximum Likelihood

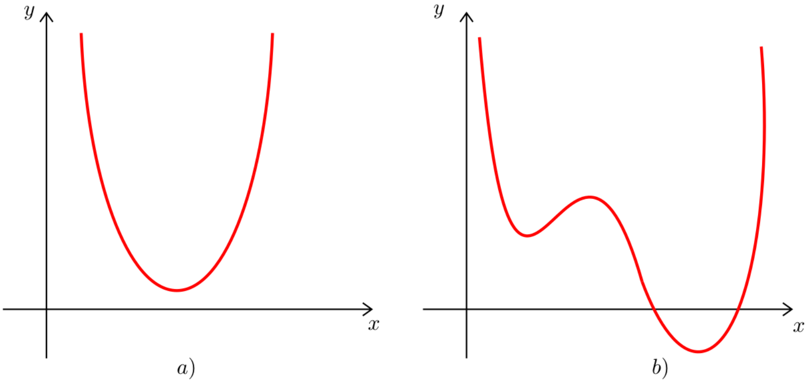
$$
\begin{aligned}
L(\boldsymbol{\theta}) &= p(\boldsymbol{y}|\boldsymbol{X};\boldsymbol{\theta}) \\
&= \prod_{i=1}^{m} p\left(y^{(i)}|\boldsymbol{x}^{(i)};\boldsymbol{\theta}\right) \\
&= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y^{(i)} - h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})\right)^2}{2\sigma^2}\right)
\end{aligned}
$$

Maximum Likelihood: Tìm $\boldsymbol{\theta}$ sao cho $L(\boldsymbol{\theta})$ cực đại

Maximum Log-Likelihood

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \log L(\boldsymbol{\theta}) \\
&= \log \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y^{(i)} - h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})\right)^2}{2\sigma^2}\right) \\
&= \sum_{i=1}^{m} \log \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y^{(i)} - h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})\right)^2}{2\sigma^2}\right) \\
&= m \log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2} \cdot \frac{1}{2\sigma^2} \sum_{i=1}^{m} \left(y^{(i)} - h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})\right)^2
\end{aligned}
$$

Tối ưu toàn cục và cục bộ



Huấn luyện mô hình

- Tìm giá trị các tham số $\boldsymbol{w}$ và b sao cho hàm chi phí $J(\boldsymbol{\theta})$ cực tiểu

$$\underset{\boldsymbol{\theta}}{\text{minimize}}\ J(\boldsymbol{\theta})$$

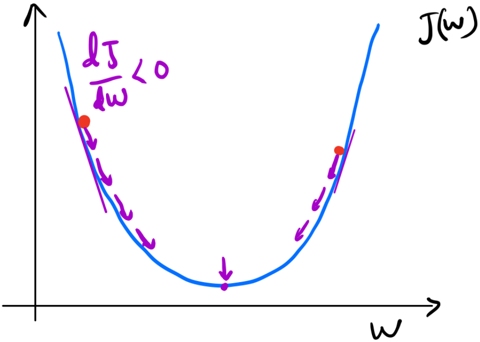
hay

$$\underset{\boldsymbol{w},b}{\text{minimize}}\ J(\boldsymbol{w}, b)$$

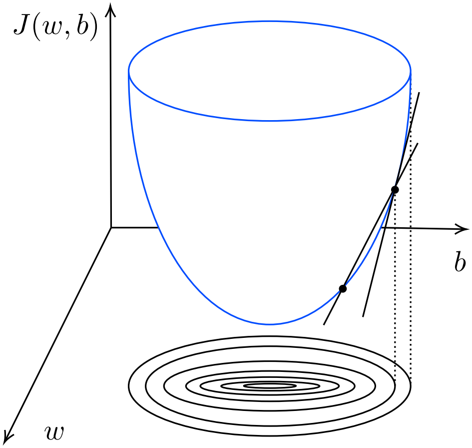
- Giải thuật huấn luyện: Giảm độ dốc (Gradient descent)

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \alpha \frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

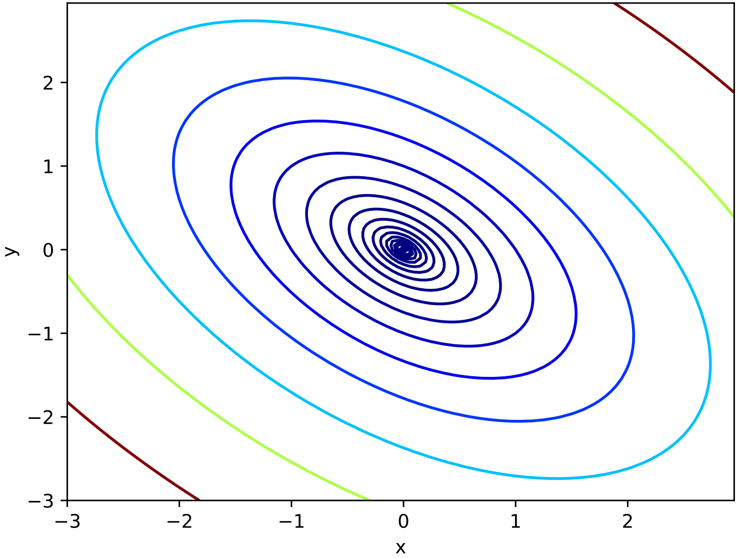
Giảm độ dốc



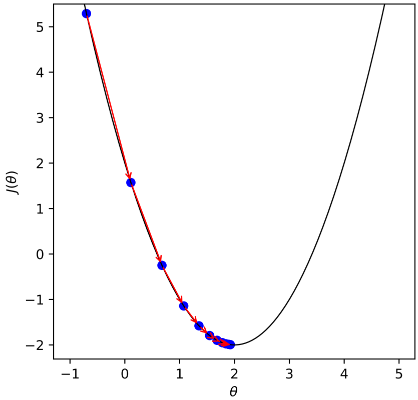
Giảm độ dốc

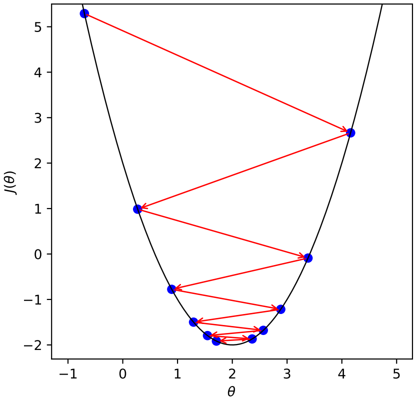


Giảm độ dốc



Giảm độ dốc





Giảm độ dốc

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \alpha \frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

Với $\boldsymbol{\theta} = \{\boldsymbol{w}, b\}$:

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \alpha \frac{\partial J(\boldsymbol{w}, b)}{\partial \boldsymbol{w}}$$

$$b \leftarrow b - \alpha \frac{\partial J(\boldsymbol{w}, b)}{\partial b}$$

Giả sử loss là bình phương sai lệch:

$$\mathcal{L}(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2 = \frac{1}{2}(h_{\boldsymbol{\theta}}(\boldsymbol{x}) - y)^2$$

Giảm độ dốc

1 mẫu huấn luyện:

$$\frac{\partial J(\boldsymbol{w}, b)}{\partial w_j} = \mathcal{L}(\hat{y}, y) \tag{1}$$

$$= \frac{\partial}{\partial w_j} \frac{1}{2} (h_{\boldsymbol{\theta}}(\boldsymbol{x}) - y)^2 \tag{2}$$

$$= 2\frac{1}{2}(h_{\boldsymbol{\theta}}(\boldsymbol{x}) - y) \frac{\partial}{\partial w_j} (h_{\boldsymbol{\theta}}(\boldsymbol{x}) - y) \tag{3}$$

$$= (h_{\boldsymbol{\theta}}(\boldsymbol{x}) - y) \frac{\partial}{\partial w_j} \left(\sum_{i=1}^{n} (w_i x_i + b) - y \right) \tag{4}$$

$$= (h_{\boldsymbol{\theta}}(\boldsymbol{x}) - y) x_j \tag{5}$$

Giảm độ dốc

m mẫu huấn luyện:

Giảm độ dốc

m mẫu huấn luyện:

$$\frac{\partial J(\boldsymbol{w}, b)}{\partial w_j} = \sum_{i=1}^{m} (h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) - y^{(i)}) x_j^{(i)} \tag{6}$$

$$\frac{\partial J(\boldsymbol{w}, b)}{\partial b} = \sum_{i=1}^{m} (h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) - y^{(i)}) \tag{7}$$

Luật cập nhật:

$$w_j \leftarrow w_j - \alpha \sum_{i=1}^{m} (h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) - y^{(i)}) x_j^{(i)} \tag{8}$$

$$b \leftarrow b - \alpha \sum_{i=1}^{m} (h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) - y^{(i)}) \tag{9}$$

Batch Gradient descent (GD)

Lặp $epoch$ lần {

Cập nhật $\boldsymbol{w}$, $(j = \overline{1, n})$:

$$w_j \leftarrow w_j - \alpha \sum_{i=1}^{m} (h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) - y^{(i)}) x_j^{(i)}$$

Cập nhật b:

$$b \leftarrow b - \alpha \sum_{i=1}^{m} (h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) - y^{(i)})$$

}

Stochastic gradient descent (SGD)

Lặp $epoch$ lần {

for i = 1 to m:

Cập nhật $\boldsymbol{w}$, $(j = \overline{1, n})$:

$$w_j \leftarrow w_j - \alpha(h_{\boldsymbol{\theta}}(\boldsymbol{x}) - y)x_j$$

Cập nhật b:

$$b \leftarrow b - \alpha(h_{\boldsymbol{\theta}}(\boldsymbol{x}) - y)$$

}

Mini-batch gradient descent

Lặp $epoch$ lần {

Lấy một khối m mẫu từ tập huấn luyện

$\{(\boldsymbol{x}^{(1)}, y^{(1)}), (\boldsymbol{x}^{(2)}, y^{(2)}), ..., (\boldsymbol{x}^{(m)}, y^{(m)})\}$

Cập nhật $\boldsymbol{w}$, $(j = \overline{1, n})$:

$$w_j \leftarrow w_j - \alpha \sum_{i=1}^{m} (h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) - y^{(i)}) x_j^{(i)}$$

Cập nhật b:

$$b \leftarrow b - \alpha \sum_{i=1}^{m} (h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) - y^{(i)})$$

}