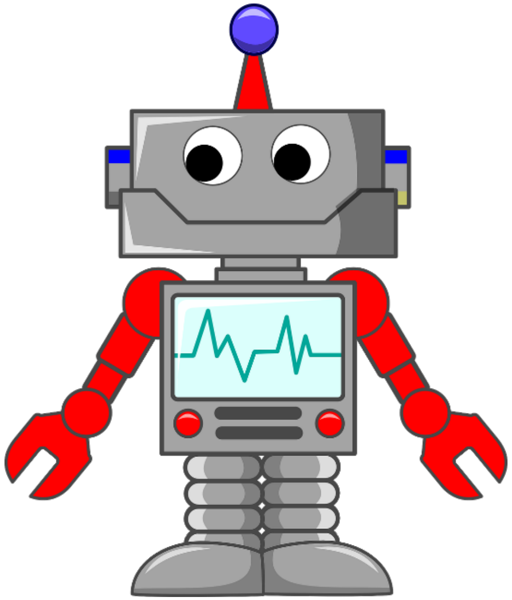


CÁC MÔ HÌNH HỌC SÂU VÀ ỨNG DỤNG

MÔ HÌNH
MÁY HỌC TỔNG QUÁT

Mô hình học tổng quát

Data x → Mô hình máy học $f_\theta(x)$ → dự đoán $\tilde{y}$ ≈ dữ liệu thực y

$$\underbrace{\tilde{y} \approx y}_{\mathcal{L}(\theta; x, y)}$$

Mô hình học tổng quát

Data x → $\boxed{\text{Mô hình máy học } f_\theta(x)}$ → $\underbrace{\text{dự đoán } \tilde{y} \approx \text{dữ liệu thực } y}_{\mathcal{L}(\theta; x, y)}$

- Huấn luyện mô hình từ tập dữ liệu huấn luyện $\{(x_i, y_i)\}_{i=1..n}$
- Hay, tìm tham số θ của mô hình f để $\tilde{y} \approx y$
- Hay, tìm tham số θ để hàm độ lỗi $\mathcal{L}(\theta; x, y)$ nhỏ nhất

Mô hình học tổng quát

Data x → **Mô hình máy học** $f_\theta(x)$ → dự đoán $\tilde{y}$ $\approx$ dữ liệu thực y

$$\underbrace{\tilde{y} \approx y}_{\mathcal{L}(\theta; x, y)}$$

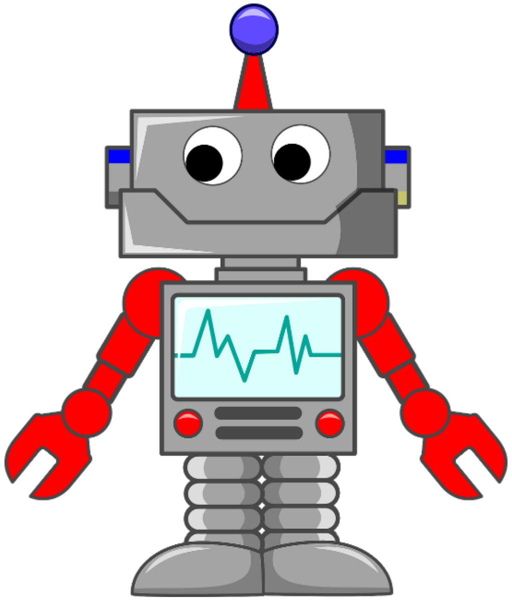
$$\begin{aligned}
&\underset{\theta}{\operatorname{argmin}}\, \mathcal{L}(\theta; x, y) \\
&\cong \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} \mathrm{D}(\tilde{y}_i, y_i) \\
&= \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} \mathrm{D}(f_\theta(x_i), y_i)
\end{aligned}$$

Thuật toán Gradient Descent

Data x ⇒ Mô hình máy học $f_\theta(x)$ ⇒ dự đoán $\tilde{y}$ $\approx$ dữ liệu thực y

$\mathcal{L}(\theta; x, y)$

- Tính đạo hàm riêng $\frac{\delta \mathcal{L}}{\delta \theta}$
- Khởi tạo: $\theta \leftarrow \theta_0$, hằng số $\alpha > 0$ và $\varepsilon > 0$ đủ nhỏ
- Lặp:
  – Cập nhật $\theta \leftarrow \theta - \alpha \frac{\delta \mathcal{L}(\theta)}{\delta \theta}$
  – Nếu $\left|\frac{\delta \mathcal{L}(\theta)}{\delta \theta}\right| < \varepsilon$: dừng lặp
- θ là tham số để $\mathcal{L}$ đạt cực tiểu

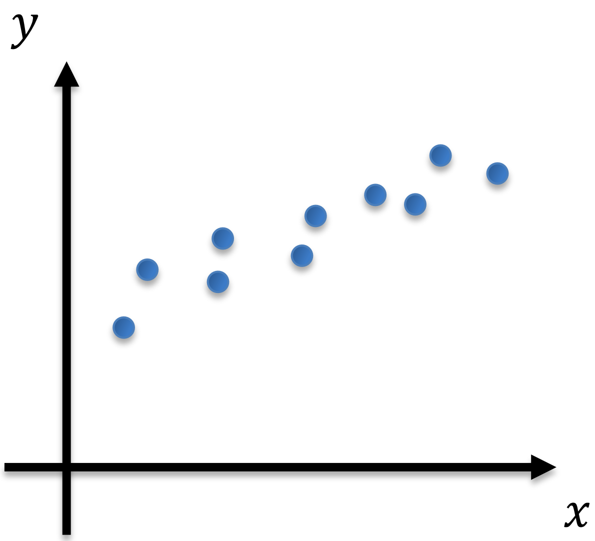


CÁC KỸ THUẬT HỌC SÂU VÀ ỨNG DỤNG

MÔ HÌNH
HỒI QUY TUYẾN TÍNH
(LINEAR REGRESSION)

Mô hình Linear Regression

Data x → **Mô hình máy học $f_\theta(x)$** → dự đoán $\underbrace{\tilde{y} \approx \text{dữ liệu thực } y}_{\mathcal{L}(\theta; x, y)}$



Chọn $f_\theta(x_i) = \theta_1 x_i + \theta_0$, khi đó ta cần tìm:

$$\underset{\theta}{\operatorname{agrmin}} \sum_{i=1}^{n} \mathrm{D}(f_\theta(x_i), y_i) \cong \underset{\theta_1,\theta_0}{\operatorname{agrmin}} \frac{1}{2n} \sum_{i=1}^{n} (\theta_1 x_i + \theta_0 - y_i)^2$$

Cần tính $\frac{\delta \mathcal{L}}{\delta \theta_1}$ và $\frac{\delta \mathcal{L}}{\delta \theta_0}$

Mô hình Linear Regression

$$\mathcal{L}(\theta_0, \theta_1) \cong \frac{1}{2n}\sum_{i=1}^{n}(\theta_1 x_i + \theta_0 - y_i)^2$$

$$\frac{\delta\mathcal{L}}{\delta\theta_0} = \frac{1}{n}\sum_{i=1}^{n}(\theta_1 x_i + \theta_0 - y_i)$$

$$\frac{\delta\mathcal{L}}{\delta\theta_1} = \frac{1}{n}\sum_{i=1}^{n}(\theta_1 x_i + \theta_0 - y_i)x_i$$

- Tính đạo hàm riêng $\frac{\delta\mathcal{L}}{\delta\theta}$
- Khởi tạo: θ_0, θ_1 ngẫu nhiên, $\alpha, \varepsilon > 0$ đủ nhỏ
- Lặp:
  - $\theta_0 \leftarrow \theta_0 - \alpha\frac{\delta\mathcal{L}}{\delta\theta_0}$
  - $\theta_1 \leftarrow \theta_1 - \alpha\frac{\delta\mathcal{L}}{\delta\theta_1}$
  - Nếu $\left|\frac{\delta\mathcal{L}}{\delta\theta_0}\right| < \varepsilon$ và $\left|\frac{\delta\mathcal{L}}{\delta\theta_1}\right| < \varepsilon$: dừng lặp
- θ là tham số để $\mathcal{L}$ đạt cực tiểu

Vector hóa công thức

Đặt $\boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$, $\bar{\mathbf{x}} = \begin{bmatrix} 1 \\ x \end{bmatrix}$, $X = \begin{bmatrix} 1 & 1 & & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}$, $Y = [y_1 \quad y_2 \quad \cdots \quad y_n]$

2×1

$f_{\boldsymbol{\theta}}(x) = \boldsymbol{\theta}^T \bar{\mathbf{x}}$

$$\nabla_\theta \mathcal{L} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial \theta_0} \\ \frac{\partial \mathcal{L}}{\partial \theta_1} \end{bmatrix} = \frac{1}{n} X(\theta^T X - Y)^T$$

$$\frac{\delta \mathcal{L}}{\delta \theta_0} = \frac{1}{n} \sum_{i=1}^{n} (\theta_1 x_i + \theta_0 - y_i) = \frac{1}{n} X_1 (\boldsymbol{\theta}^T X - Y)^T$$

$$\frac{\delta \mathcal{L}}{\delta \theta_1} = \frac{1}{n} \sum_{i=1}^{n} (\theta_1 x_i + \theta_0 - y_i) x_i = \frac{1}{n} X_2 (\boldsymbol{\theta}^T X - Y)^T$$

Mô hình Linear Regression đa biến

Dữ liệu đầu vào được ký hiệu: $\mathbf{x}^{(i)} \in \mathbb{R}^m$, dữ liệu đầu ra $y^{(i)} \in \mathbb{R}$, khi đó:

$$\mathcal{L}(\theta_0, \boldsymbol{\theta}_1) \cong \frac{1}{2n}\sum_{i=1}^{n}\left(\boldsymbol{\theta}_1^T\mathbf{x}^{(i)} + \theta_0 - y^{(i)}\right)^2$$

$$\frac{\delta\mathcal{L}}{\delta\theta_0} = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\theta}_1^T\mathbf{x}^{(i)} + \theta_0 - y^{(i)}\right)$$

$$\frac{\delta\mathcal{L}}{\delta\theta_{1,k}} = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\theta}_1^T\mathbf{x}^{(i)} + \theta_0 - y^{(i)}\right)x_k^{(i)}$$

- Tính đạo hàm riêng $\frac{\delta\mathcal{L}}{\delta\theta}$
- Khởi tạo: $\theta_0, \boldsymbol{\theta}_1$ ngẫu nhiên, $\alpha, \varepsilon > 0$ đủ nhỏ
- Lặp:
  - $\theta_0 \leftarrow \theta_0 - \alpha\frac{\delta\mathcal{L}}{\delta\theta_0}$
  - $\theta_{1,k} \leftarrow \theta_1 - \alpha\frac{\delta\mathcal{L}}{\delta\theta_{1,k}}$, với $k = 1..m$
  - Nếu $\left|\frac{\delta\mathcal{L}}{\delta\theta_0}\right| < \varepsilon$ và $\left|\frac{\delta\mathcal{L}}{\delta\theta_{1,k}}\right| < \varepsilon$: dừng lặp
- $\theta_0, \boldsymbol{\theta}_1$ là tham số để $\mathcal{L}$ đạt cực tiểu

Vector hóa công thức

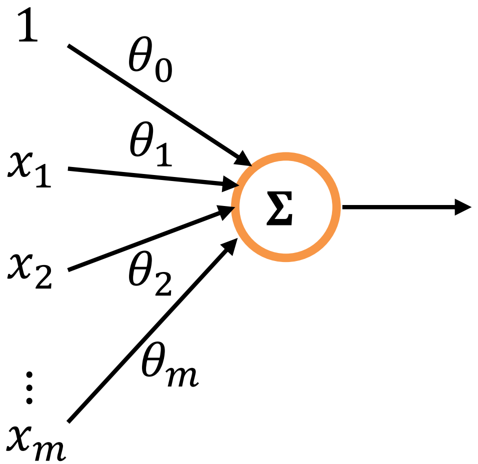
Đặt $\boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \boldsymbol{\theta}_1 \end{bmatrix}$, $X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \dots & \mathbf{x}^{(n)} \end{bmatrix}$, $Y = \begin{bmatrix} y^{(1)} & y^{(2)} & \dots & y^{(n)} \end{bmatrix}$

$$\frac{\delta \mathcal{L}}{\delta \theta_0} = \frac{1}{n} \sum_{i=1}^{n} \left(\boldsymbol{\theta}_1^T \mathbf{x}^{(i)} + \theta_0 - y^{(i)} \right)$$

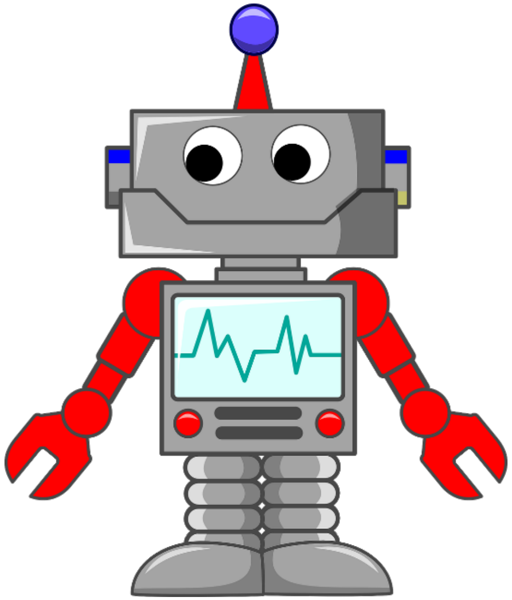
$$\frac{\delta \mathcal{L}}{\delta \theta_{1,k}} = \frac{1}{n} \sum_{i=1}^{n} \left(\boldsymbol{\theta}_1^T \mathbf{x}^{(i)} + \theta_0 - y^{(i)} \right) x_k^{(i)}$$

Dạng đồ thị của Linear Reg.

Với $\mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$



$$\tilde{y} = f_{\boldsymbol{\theta}}(\mathbf{x}) = \boldsymbol{\theta}^T \mathbf{x}$$



CÁC KỸ THUẬT HỌC SÂU VÀ ỨNG DỤNG

MÔ HÌNH HỒI QUY LUẬN LÝ (LOGISTIC REGRESSION)

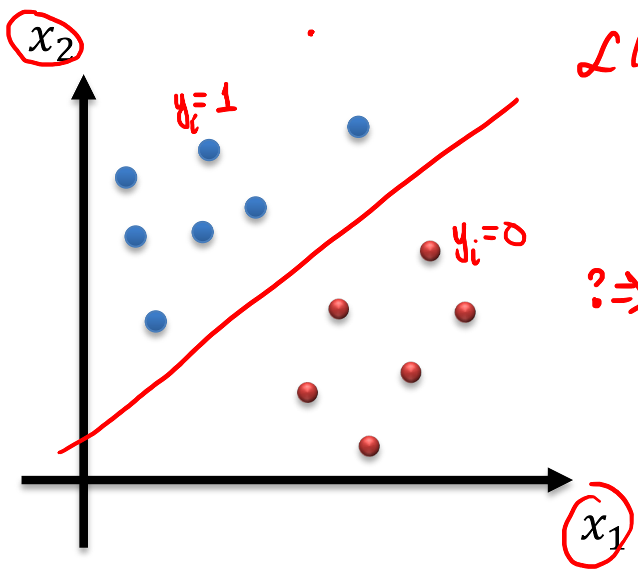
Đúng vs Sai
1 0
phân lớp nhị phân

Mô hình Logistic Regression

Data $\mathbf{x}$ → Mô hình máy học $f_\theta(x)$ → dự đoán $\tilde{y}$ ≈ dữ liệu thực y

$$\underbrace{\tilde{y} \approx y}_{\mathcal{L}(\theta; x, y)}$$

$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $y \in [0,1]$



$$\mathcal{L}(\theta) = \underset{\boldsymbol{\theta}}{\operatorname{agrmin}} \sum_{i=1}^{n} \mathrm{D}\big(\underbrace{f_{\boldsymbol{\theta}}(\mathbf{x}^{(i)})}_{\tilde{y}}, y^{(i)}\big)$$

$\tilde{y} \simeq y$

$$\cong \underset{\boldsymbol{\theta}}{\operatorname{agrmin}} \sum_{i=1}^{n} (f_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}) - y^{(i)})^2$$

$? \Rightarrow y = 0$

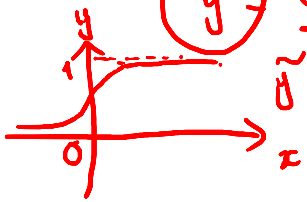
phụ thuộc $\tilde{y}$

$y^{(i)} \in [0,1]$

$\tilde{y} = \theta^T . \bar{x} \in [-\infty, \infty]$

sigmoid

$\sigma(x) = \dfrac{1}{1+e^{-x}}$

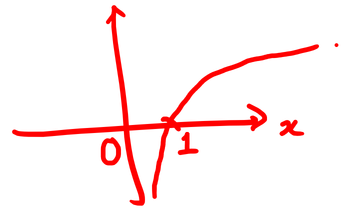


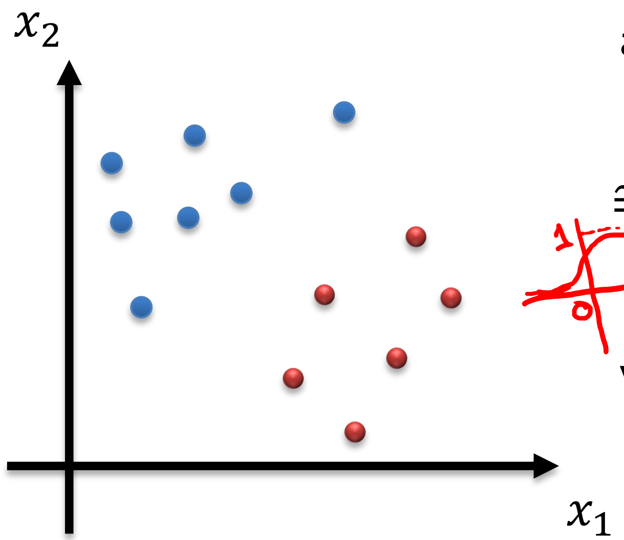
$\tilde{y} \to [0,1]$

Mô hình Logistic Regression

Data $\mathbf{x}$ → Mô hình máy học $f_\theta(x)$ → dự đoán $\tilde{y}$ $\approx$ dữ liệu thực y

$$\underbrace{\tilde{y} \approx y}_{\mathcal{L}(\theta; x, y)}$$

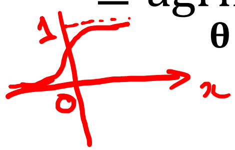




$$\underset{\boldsymbol{\theta}}{\operatorname{agrmin}} \sum_{i=1}^{n} \mathrm{D}\left(f_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}), y^{(i)}\right)$$

1 . log 0 → -∞ 0

$$\cong \underset{\boldsymbol{\theta}}{\operatorname{agrmin}} - \sum_{i=1}^{n} y^{(i)} \log \underbrace{f_{\boldsymbol{\theta}}(\mathbf{x}^{(i)})}_{\tilde{y}} + (1 - y^{(i)}) \log(1 - \underbrace{f_{\boldsymbol{\theta}}(\mathbf{x}^{(i)})}_{\tilde{y}}))$$



Với $f_{\boldsymbol{\theta}}(\mathbf{x}) = \dfrac{1}{1+e^{-(\theta^T \bar{\mathbf{x}})}}$

$-\infty, \infty \Rightarrow 0 \to 1$

Đoán đúng : $y = \tilde{y} = 0 \to \mathcal{L} = 0$

$y = \tilde{y} = 1 \to \mathcal{L} = 0$

Đoán sai : $1 = y \neq \tilde{y} = 0 \to \mathcal{L} = +\infty$

↓ bán nhà

Mô hình Logistic Regression

$\frac{\partial \mathcal{L}}{\partial \theta_k} \rightarrow$ gradient descent

$f_\theta(x^{(i)}) = \sigma(\theta^T \bar{x}^{(i)})$

$\sigma(x)$

$\sigma'(x) = \sigma(x).(1 - \sigma(x))$

$$\mathcal{L}(\boldsymbol{\theta}) \cong -\sum_{i=1}^{n} y^{(i)} \log f_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}) + (1 - y^{(i)}) \log(1 - f_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}))$$

$$\frac{\delta \mathcal{L}}{\delta \theta_k} = -y^{(i)} \sum_{i=1}^{n} \frac{1}{\sigma(\boldsymbol{\theta}^T \bar{\mathbf{x}}^{(i)})} \frac{\delta \sigma(\boldsymbol{\theta}^T \bar{\mathbf{x}}^{(i)})}{\delta \theta_k} + (1 - y^{(i)}) \sum_{i=1}^{n} \frac{1}{1 - \sigma(\boldsymbol{\theta}^T \bar{\mathbf{x}}_i)} \frac{\delta \sigma(\boldsymbol{\theta}^T \bar{\mathbf{x}}^{(i)})}{\delta \theta_k}$$

f.

$$= -y^{(i)} \sum_{i=1}^{n} \frac{1}{\cancel{\sigma(\boldsymbol{\theta}^T \bar{\mathbf{x}}^{(i)})}} \cancel{\sigma(\boldsymbol{\theta}^T \bar{\mathbf{x}}^{(i)})} \left(1 - \sigma\left(\boldsymbol{\theta}^T \bar{\mathbf{x}}^{(i)}\right)\right) \frac{\delta\left(\boldsymbol{\theta}^T \bar{\mathbf{x}}^{(i)}\right)}{\delta \theta_k}$$

$$+(1 - y^{(i)}) \sum_{i=1}^{n} \frac{1}{\cancel{1 - \sigma(\boldsymbol{\theta}^T \bar{\mathbf{x}}^{(i)})}} \sigma(\boldsymbol{\theta}^T \bar{\mathbf{x}}^{(i)}) \cancel{(1 - \sigma(\boldsymbol{\theta}^T \bar{\mathbf{x}}^{(i)}))} \frac{\delta(\boldsymbol{\theta}^T \bar{\mathbf{x}}^{(i)})}{\delta \theta_k}$$

Mô hình Logistic Regression

$$\frac{\delta \mathcal{L}}{\delta \theta_k} = -y^{(i)} \sum_{i=1}^{n} \left(1 - \cancel{\sigma\left(\boldsymbol{\theta}^T \overline{\mathbf{x}}^{(i)}\right)}\right) \frac{\delta\left(\boldsymbol{\theta}^T \overline{\mathbf{x}}^{(i)}\right)}{\delta \theta_k} + (1 - \cancel{y^{(i)}}) \sum_{i=1}^{n} \sigma(\boldsymbol{\theta}^T \overline{\mathbf{x}}^{(i)}) \frac{\delta(\boldsymbol{\theta}^T \overline{\mathbf{x}}^{(i)})}{\delta \theta_k}$$

$$= -\sum_{i=1}^{n} y^{(i)} \frac{\delta\left(\boldsymbol{\theta}^T \overline{\mathbf{x}}^{(i)}\right)}{\delta \theta_k} + \sum_{i=1}^{n} \sigma(\boldsymbol{\theta}^T \overline{\mathbf{x}}^{(i)}) \frac{\delta(\boldsymbol{\theta}^T \overline{\mathbf{x}}^{(i)})}{\delta \theta_k}$$

$$= \sum_{i=1}^{n} (\sigma\left(\boldsymbol{\theta}^T \overline{\mathbf{x}}^{(i)}\right) - y^{(i)}) \frac{\delta(\boldsymbol{\theta}^T \overline{\mathbf{x}}^{(i)})}{\delta \theta_k} = \sum_{i=1}^{n} (\sigma\left(\boldsymbol{\theta}^T \overline{\mathbf{x}}^{(i)}\right) - y^{(i)}) \bar{x}_k^{(i)}$$

quả lại $\bar{x}_k^{(i)}$.

Vector hóa công thức

$$Y = \left[y^{(1)} \; y^{(2)} \; \dots \; y^{(n)} \right]$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_m \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 1 & & 1 \\ x_1^{(1)} & x_1^{(2)} & & x_1^{(n)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(n)} \\ \vdots & \vdots & & \vdots \\ x_m^{(1)} & x_m^{(2)} & & x_m^{(n)} \end{bmatrix}$$

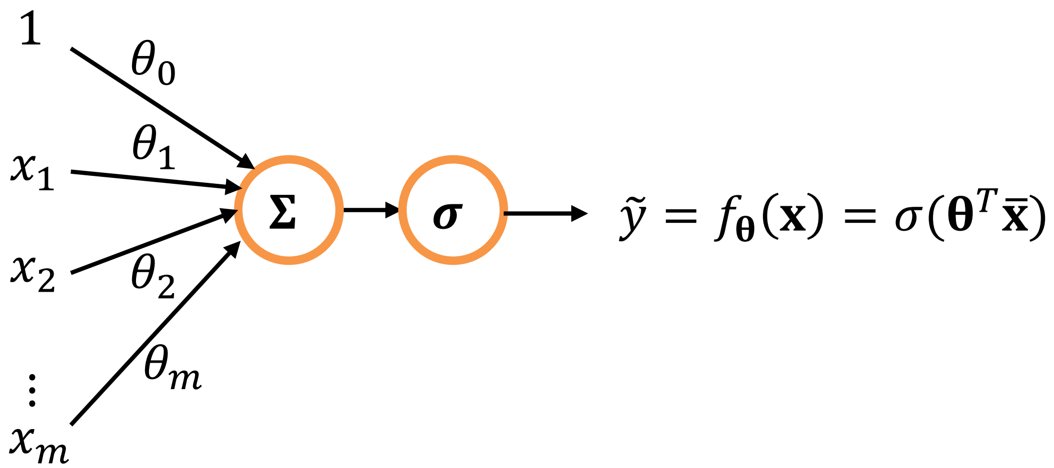
$$\frac{\delta \mathcal{L}}{\delta \theta_k} = \sum_{i=1}^{n} \left(\sigma\left(\boldsymbol{\theta}^T \overline{\mathbf{x}}^{(i)}\right) - y^{(i)}\right) \bar{x}_k^{(i)}$$

$$\equiv X\left(\theta^T X - Y\right)^T \rightarrow \text{vector}$$

$$\nabla_\theta \mathcal{L} = X\left(\theta^T X - Y\right)^T$$

Dạng đồ thị của Logistic Reg.

Với $\bar{\mathbf{x}} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$ và $\sigma(x) = \frac{1}{1+e^{-x}}$



Dạng đồ thị của Softmax Reg.