



# HUST

**ĐẠI HỌC BÁCH KHOA HÀ NỘI**  
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.





**ĐẠI HỌC  
BÁCH KHOA HÀ NỘI**  
HANOI UNIVERSITY  
OF SCIENCE AND TECHNOLOGY

# Linear Regression

# Logistic Regression

ONE LOVE. ONE FUTURE.

- Linear Regression
- Logistic Regression, Softmax Regression
- Ứng dụng trong bài toán thực tế

- Linear Regression phân tích quan hệ giữa biến phụ thuộc với một hoặc nhiều biến độc lập. Mô hình hóa sử dụng hàm tuyến tính

$$y \approx \hat{y} = f(\mathbf{x}) = w_1x_1 + w_2x_2 + w_3x_3 = \mathbf{x}^T \mathbf{w}$$

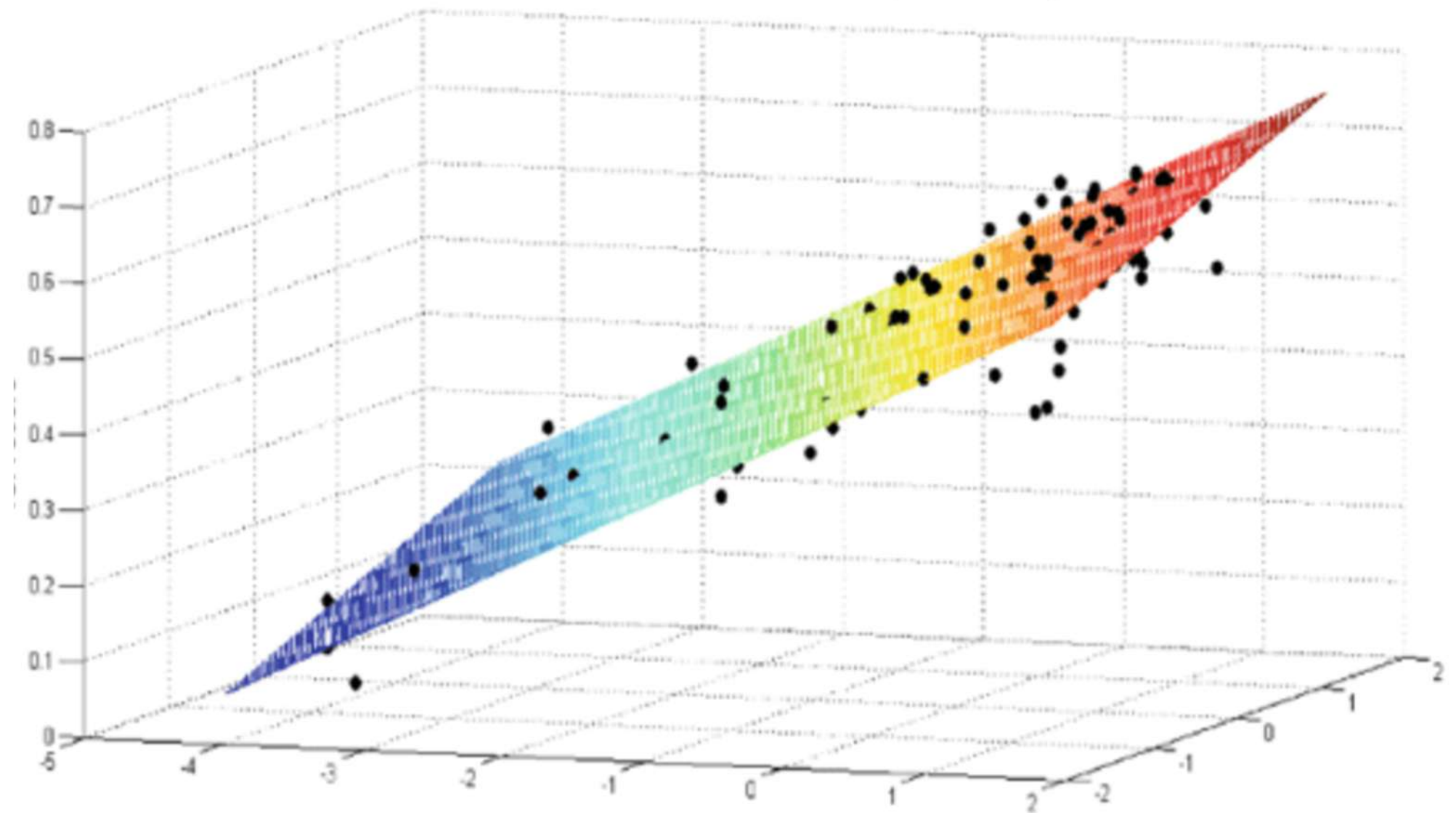
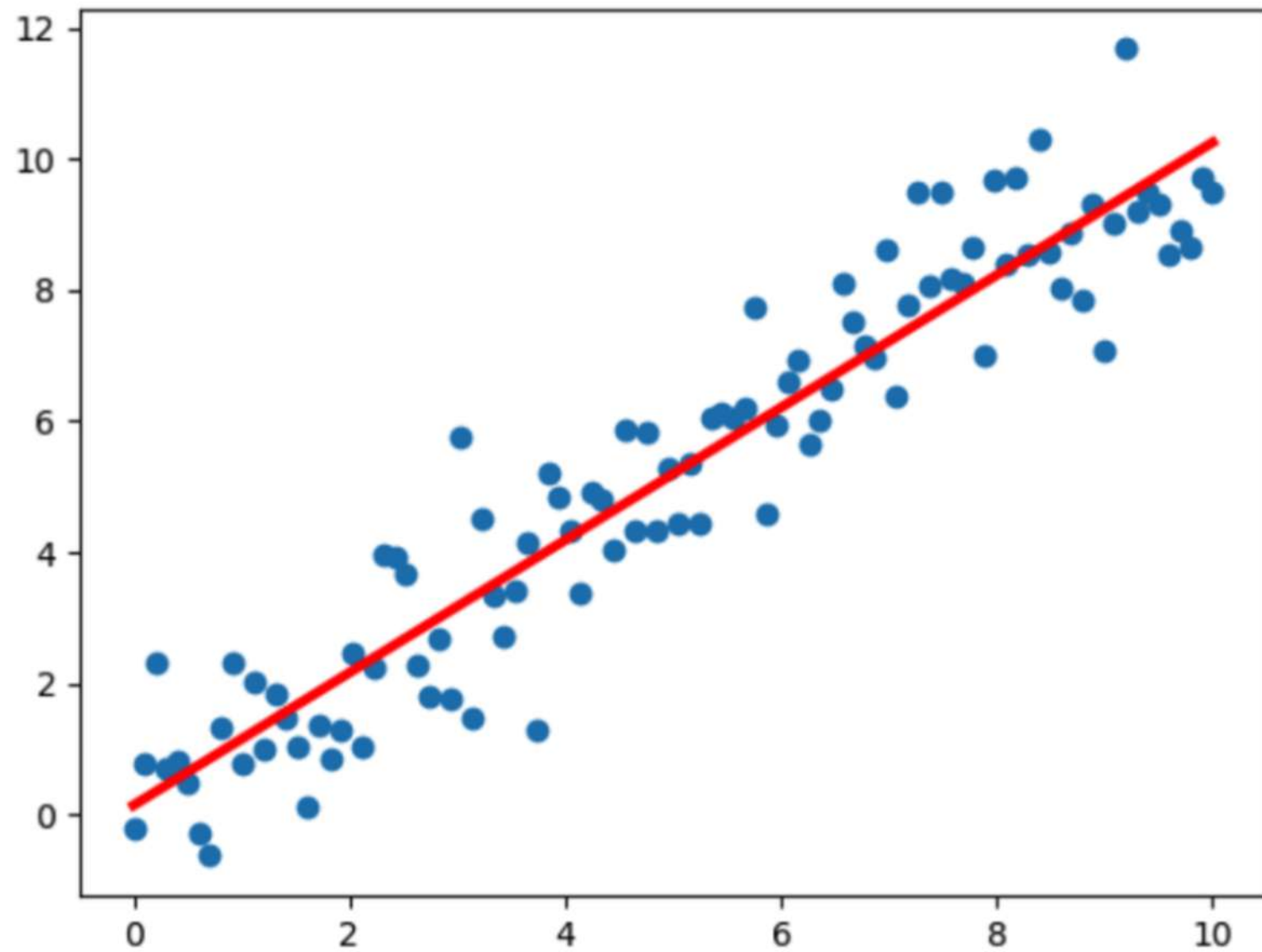
- Hàm mất mát chính là trung bình sai số của tất cả các điểm dữ liệu

$$\mathcal{L}(\mathbf{w}) = \frac{1}{2N} \sum_{i=1}^N (y_i - \mathbf{x}_i^T \mathbf{w})^2 = \frac{1}{2N} \left\| \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} - \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix} \mathbf{w} \right\|_2^2 = \frac{1}{2N} \|\mathbf{y} - \mathbf{X}^T \mathbf{w}\|_2^2$$

$$\frac{\nabla \mathcal{L}(\mathbf{w})}{\nabla \mathbf{w}} = \frac{1}{N} \mathbf{X}(\mathbf{X}^T \mathbf{w} - \mathbf{y}) = \mathbf{0} \Leftrightarrow \mathbf{X} \mathbf{X}^T \mathbf{w} = \mathbf{X} \mathbf{y}$$



# Linear Regression



- Linear Regression dùng trong các bài toán hồi quy, khi biến mục tiêu là số thực liên tục và mối quan hệ giữa các biến gần như tuyến tính
- Ưu điểm của Linear Regression:
  - Đơn giản, độ phức tạp thấp, dễ triển khai
  - Dễ diễn giải và phân tích
- Nhược điểm của Linear Regression:
  - Rất nhạy với nhiễu
  - Không biểu diễn được các mối quan hệ phức tạp

- Huber Regression:

$$L_{\delta}(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & \text{for } |y - f(x)| \leq \delta, \\ \delta(|y - f(x)| - \frac{1}{2}\delta), & \text{otherwise.} \end{cases}$$

- Lasso Regression:

$$\mathcal{L}_1(\mathbf{w}) = \frac{1}{2N}(\|\mathbf{y} - \mathbf{X}^T \mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1)$$

- Ridge Regression:

$$\mathcal{L}_2(\mathbf{w}) = \frac{1}{2N}(\|\mathbf{y} - \mathbf{X}^T \mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_2^2)$$

- Polynomial Regression:



- Logistic Regression là mô hình dùng để phân loại nhị phân

$$p(y_i|\mathbf{x}_i; \mathbf{w}) = z_i^{y_i} (1 - z_i)^{1-y_i}$$

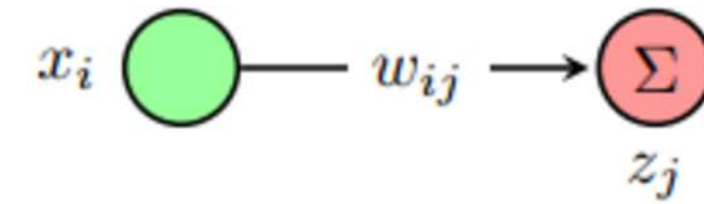
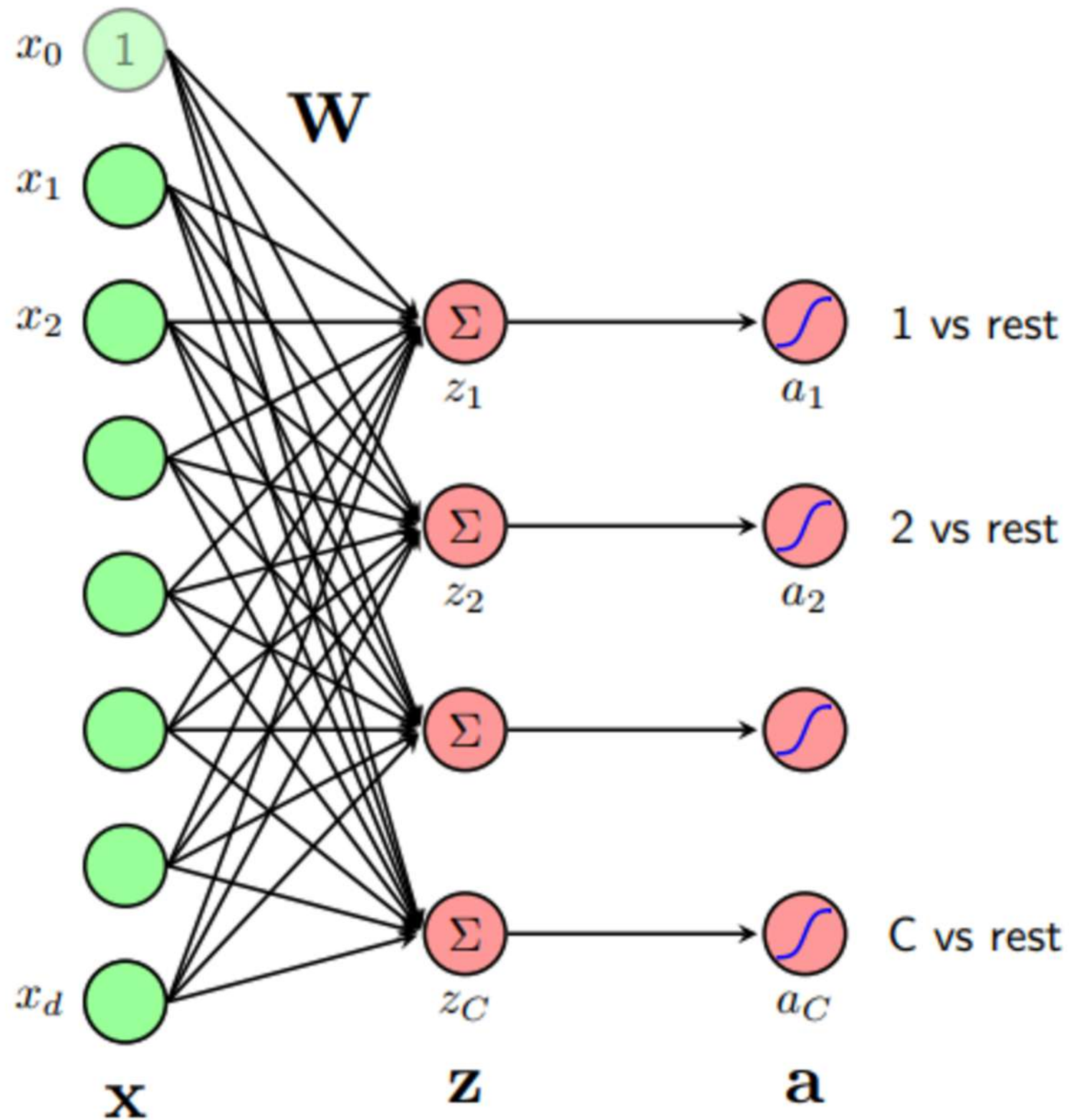
$$p(\mathbf{y}|\mathbf{X}; \mathbf{w}) = \prod_{i=1}^N p(y_i|\mathbf{x}_i; \mathbf{w}) = \prod_{i=1}^N z_i^{y_i} (1 - z_i)^{1-y_i}$$

$$J(\mathbf{w}) = -\frac{1}{N} \log p(\mathbf{y}|\mathbf{X}; \mathbf{w}) = -\frac{1}{N} \sum_{i=1}^N (y_i \log z_i + (1 - y_i) \log(1 - z_i))$$



- Logistic Regression dùng chủ yếu trong các bài toán phân loại nhị phân, phân loại đa lớp, tính xác suất các sự kiện, dùng để phân tích ảnh hưởng của từng biến đặc trưng đến xác suất
- Ưu điểm của Logistic Regression:
  - Đơn giản, dễ triển khai
  - Có thể mở rộng sang bài toán phân loại đa lớp
  - Ngưỡng ra quyết định linh hoạt
- Nhược điểm của Linear Regression:
  - Không phù hợp với dữ liệu phi tuyến do đường ranh giới tạo bởi Logistic Regression là một siêu phẳng

# Softmax Regression



$w_{0j}$ : biases, don't forget!

$d$ : data dimension

$C$ : number of classes

$$\mathbf{x} \in \mathbb{R}^{d+1}$$

$$\mathbf{W} \in \mathbb{R}^{(d+1) \times C}$$

$$z_i = \mathbf{w}_i^T \mathbf{x}$$

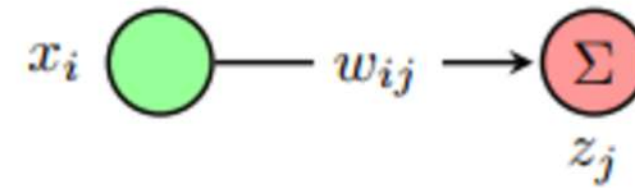
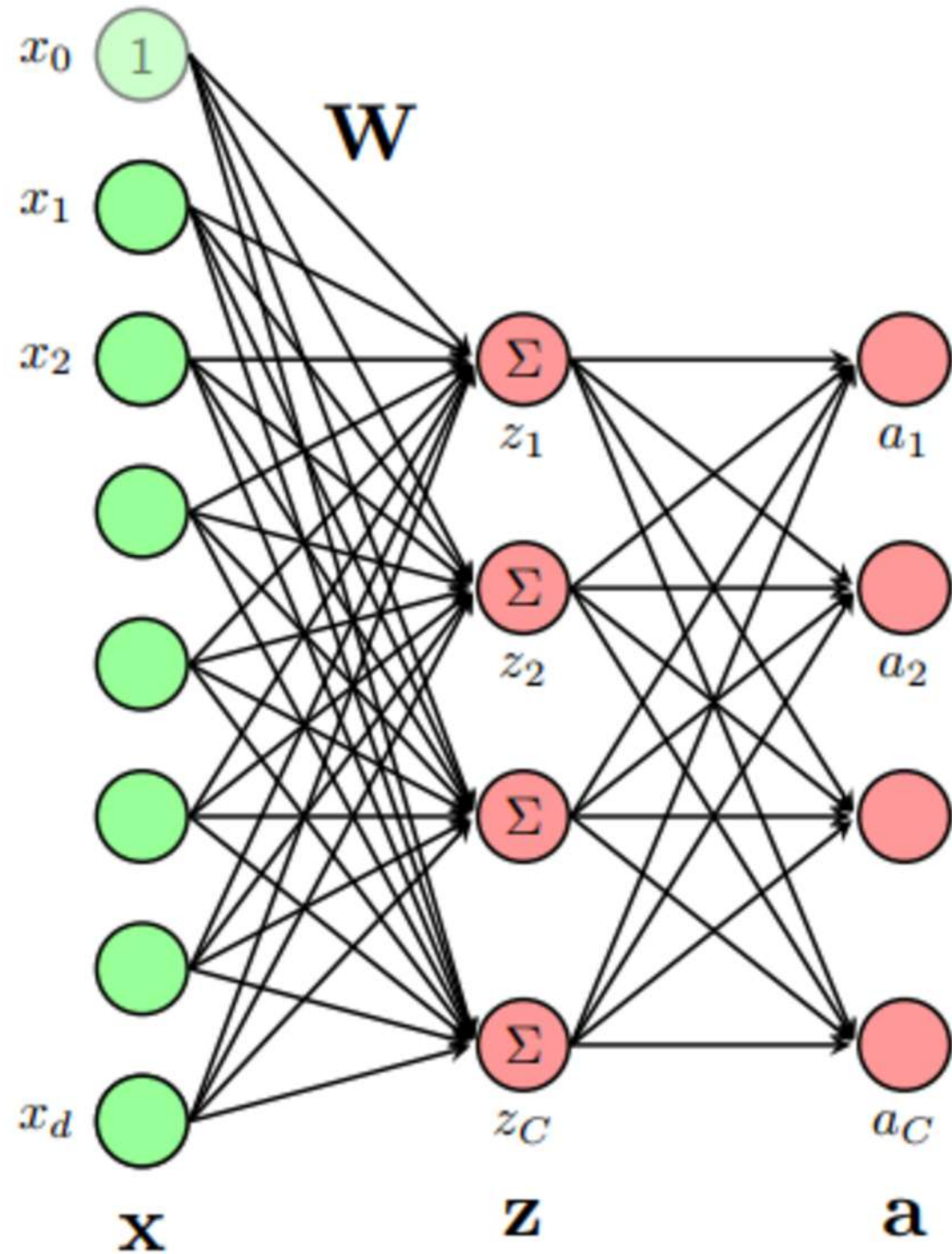
$$\mathbf{z} = \mathbf{W}^T \mathbf{x} \in \mathbb{R}^C$$

$$a_i = \text{sigmoid}(z_i) \in \mathbb{R}$$

$$0 < a_i < 1$$



# Softmax Regression



$w_{0j}$ : biases, don't forget!

$d$ : data dimension

$C$ : number of classes

$$\mathbf{x} \in \mathbb{R}^{d+1}$$

$$\mathbf{W} \in \mathbb{R}^{(d+1) \times C}$$

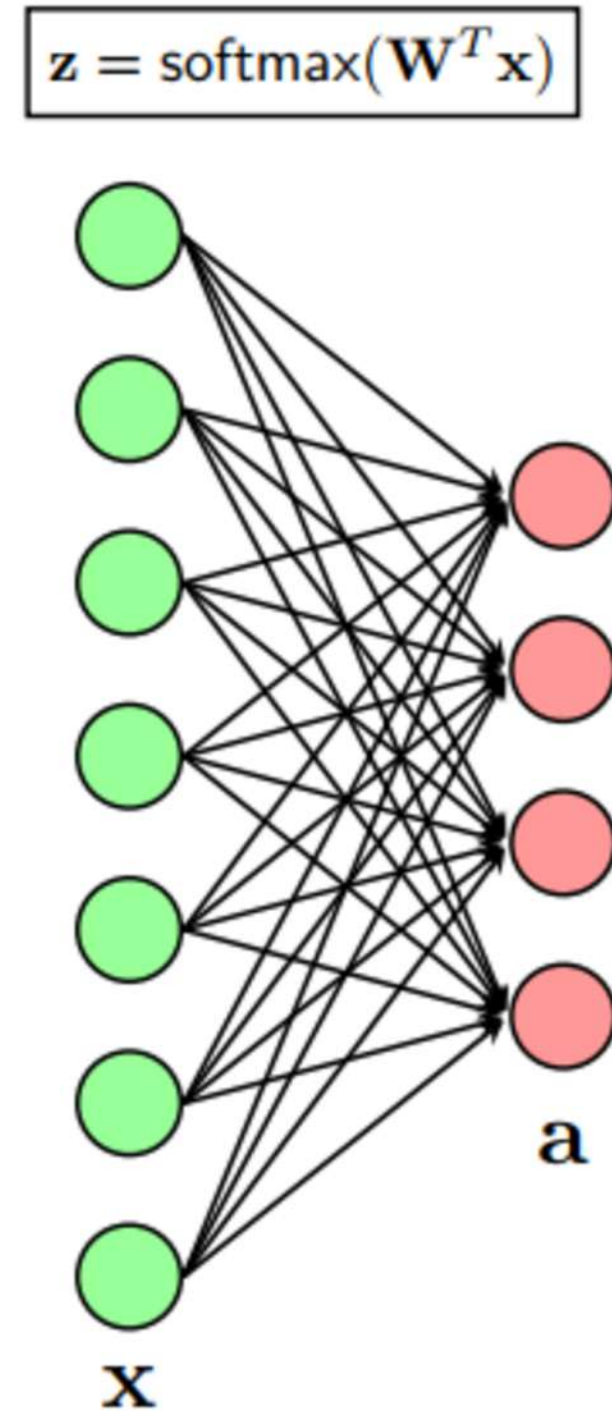
$$z_i = \mathbf{w}_i^T \mathbf{x}$$

$$\mathbf{z} = \mathbf{W}^T \mathbf{x} \in \mathbb{R}^C$$

$$\mathbf{a} = \text{softmax}(\mathbf{z}) \in \mathbb{R}^C$$

$$a_i > 0, \quad \sum_{i=1}^C a_i = 1$$

short form  $\longrightarrow$



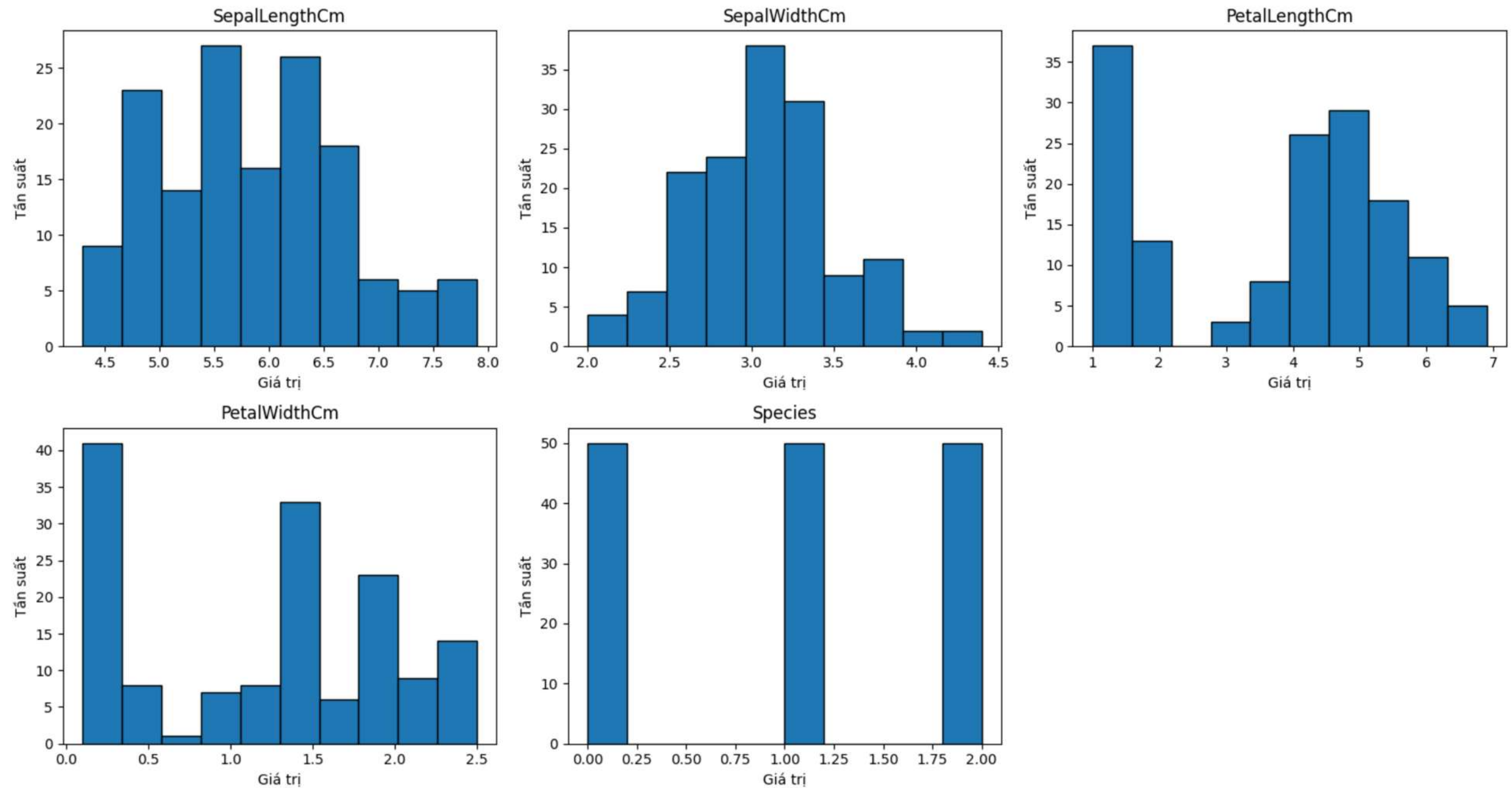
- Softmax Regression là mở rộng của Logistic Regression với số lớp phân loại lớn hơn 2, nên có tên gọi khác là Multinomial Logistic Regression
- Tương tự Logistic Regression, ranh giới tạo bởi Softmax Regression cũng là một mặt tuyến tính
- Softmax Regression là một trong hai mô hình phân loại phổ biến nhất



- Ưu điểm của Softmax Regression:
  - Phù hợp với bài toán phân loại đa lớp
  - Diễn giải xác suất của các kết quả hoặc lớp cụ thể, thể hiện mức độ quan trọng của biến đặc trưng
- Nhược điểm của Softmax Regression:
  - Phụ thuộc vào giả định tuyến tính
  - Không lý tưởng cho dữ liệu mất cân bằng lớp
  - Nhạy cảm với nhiễu

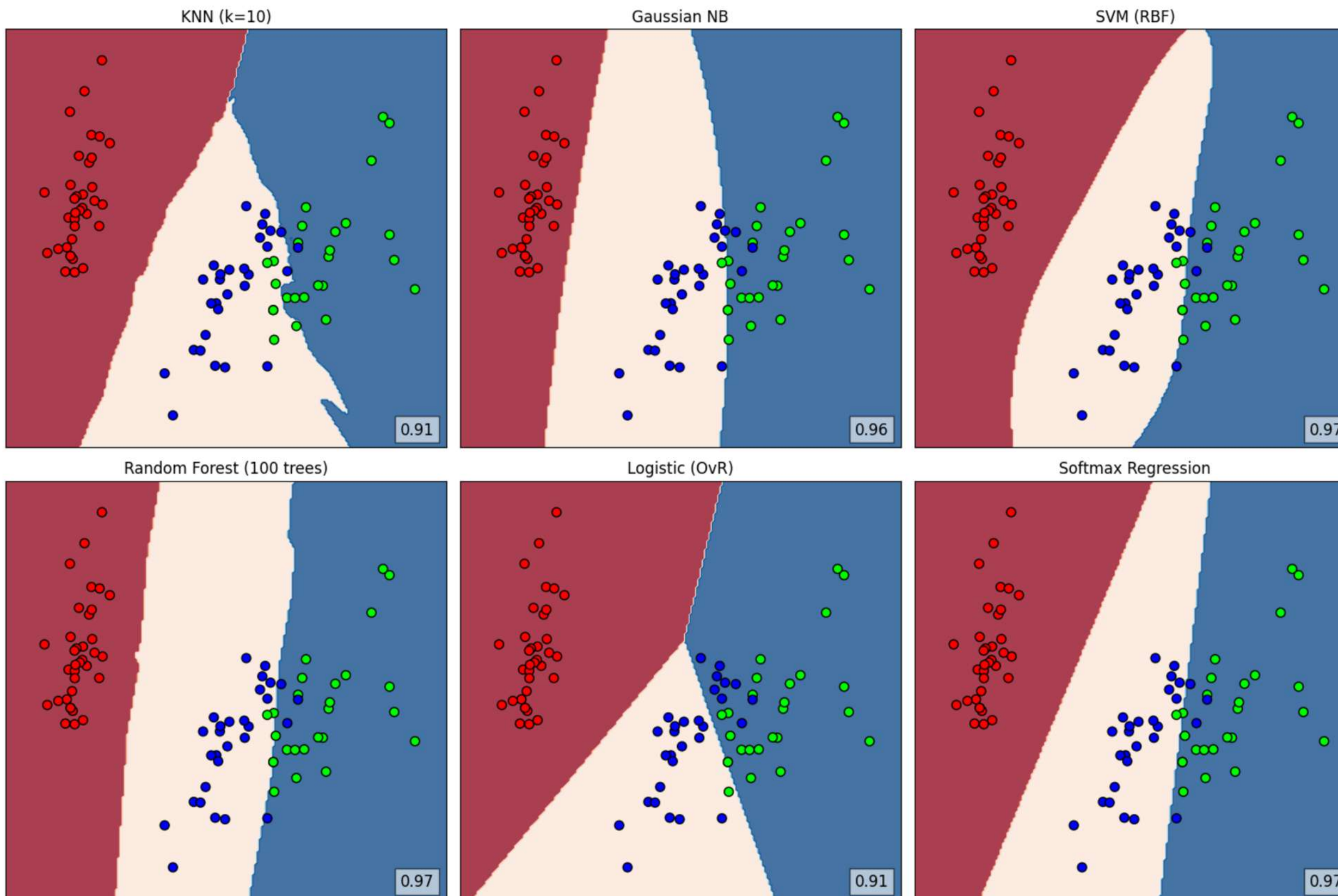
# Ứng dụng trong bài toán thực tế

- Bộ dữ liệu: Iris Classification
  - 150 mẫu
  - 3 nhãn

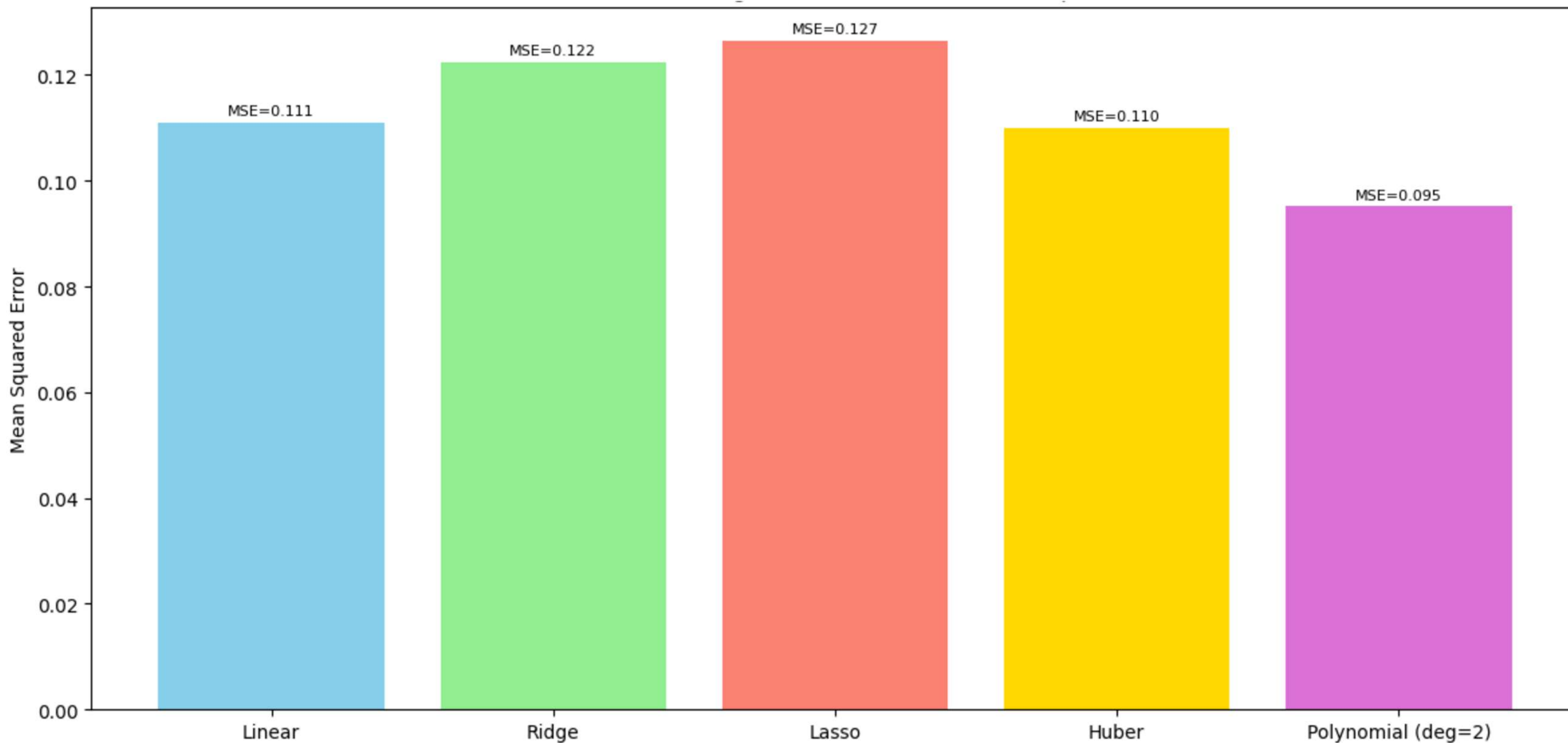




# Ứng dụng trong bài toán phân loại



# Ứng dụng trong bài toán hồi quy





- So sánh trọng số:
  - Linear: [ 0.42194898 0.4257896 -0.86767242]
  - Ridge: [ 0.34209797 0.24829824 -0.61902107]
  - Lasso: [ 0.31782012 0.17989185 -0.53311319]
  - Huber: [ 0.39419907 0.37665535 -0.804503 ]
  - Polynomial (deg=2): [ 0.31238462 0.14926653  
-0.36108044 -0.0646215 -0.27633583 0.09139712  
0.03691683 0.40121218 -0.0307105 ]

A large graphic on the left side of the slide. It features a dark blue background with a pattern of red dots of varying sizes arranged in concentric, slightly irregular circles, creating a sense of depth and movement. The word "HUST" is centered within this graphic in a white, bold, sans-serif font.

# HUST

# THANK YOU !



[hust.edu.vn](http://hust.edu.vn)



[fb.com/dhbkhn](https://fb.com/dhbkhn)