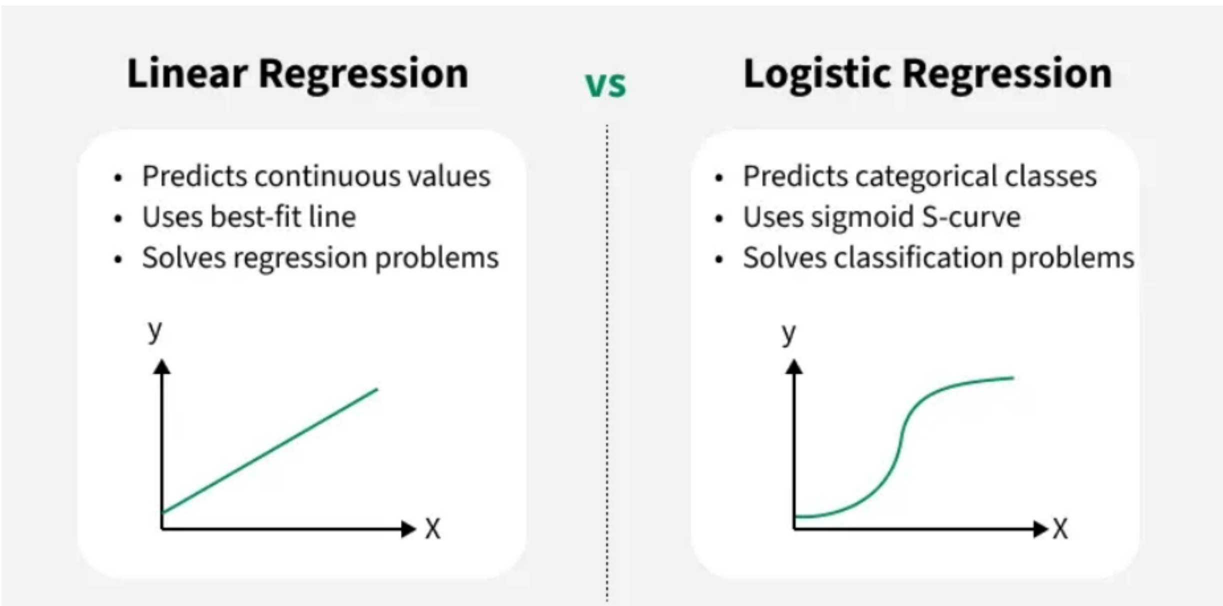
1. Lời tựa

MD06w01 giới thiệu về sự chuyển đổi từ mô hình Hồi quy Tuyến tính (Linear Regression) sang Hồi quy Logistic (Logistic Regression). Hồi quy Logistic là thuật toán nền tảng cho các bài toán phân loại nhị phân.

Sự chuyển đổi này nhằm giải quyết hạn chế của Hồi quy Tuyến tính, vốn chỉ phù hợp để dự đoán **các giá trị liên tục** trong khoảng vô hạn (-∞, +∞). Do đó, mô hình tuyến tính trở nên không phù hợp và kém hiệu quả khi áp dụng vào các tác vụ phân loại đòi hỏi đầu ra là **các nhãn rời rạc** (ví dụ: dự đoán 0 hoặc 1, Có hoặc Không). Hồi quy Logistic khắc phục điều này bằng cách sử dụng **hàm Sigmoid** để ánh xạ đầu ra về phạm vi xác suất (0, 1).



2. Hàm Sigmoid và xây dựng hàm mất mát

2.1: Tại sao sử dụng hàm sigmoid:

Khi chuyển qua bài toán phân loại, chúng ta cần chuyển đổi giá trị liên tục sang giá trị xác xuất. Từ mô hình toán học dựa trên nguyên tắc tự nhiên **Logistic Growth,** chúng ta đề xuất công thức hàm Sigmoid. Hàm Sigmoid nhận bất kỳ giá trị đầu vào tuyến tính(z) trong khoảng (-∞, +∞) và ánh xạ nó vào miền giá trị (0, 1).

z -> -∞ => e^(-z) -> +∞ => Sigmoid(z) = 1/(1 + e^(-z) -> 0

Z -> +∞ => e^(-z) -> -∞ => Sigmoid(z) = 1/(1 + e^(-z) -> 1

Sigmoid function

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

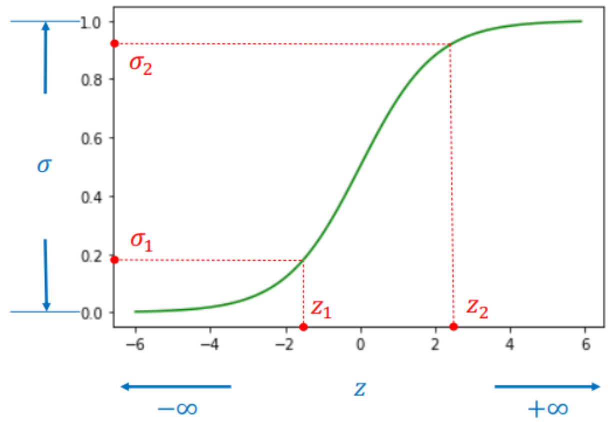
$$z \in (-\infty \quad +\infty)$$

$$\sigma(z) \in (0 \quad 1)$$

Property

$$\forall z_1 z_2 \in [a \quad b] \text{ and } z_1 \le z_2$$

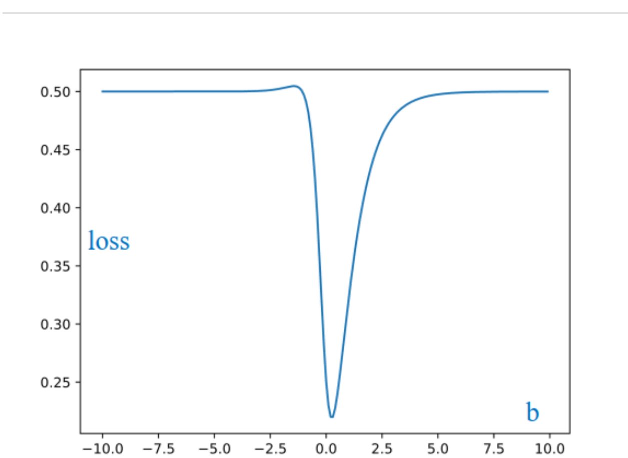
$$\rightarrow \sigma(z_1) \le z(u_1)$$



2.2. Xây dựng hàm mất mát

Việc lựa chọn hàm mất mát phù hợp là yếu tố quyết định đến hiệu suất và sự ổn định của Hồi quy Logistic. Trong tuần này, chúng ta tập trung thử nghiệm 2 loại hàm mất mát: MSE (Mean Square Error) và BCE (Binary Cross – Entroy) để đánh giá và lựa chọn cho thuật toán này.

2.2.1: MSE (Mean Square Error)



Mean Squared Error

Model and Loss

$$z = \boldsymbol{\theta}^T \boldsymbol{x} = \boldsymbol{x}^T \boldsymbol{\theta}$$

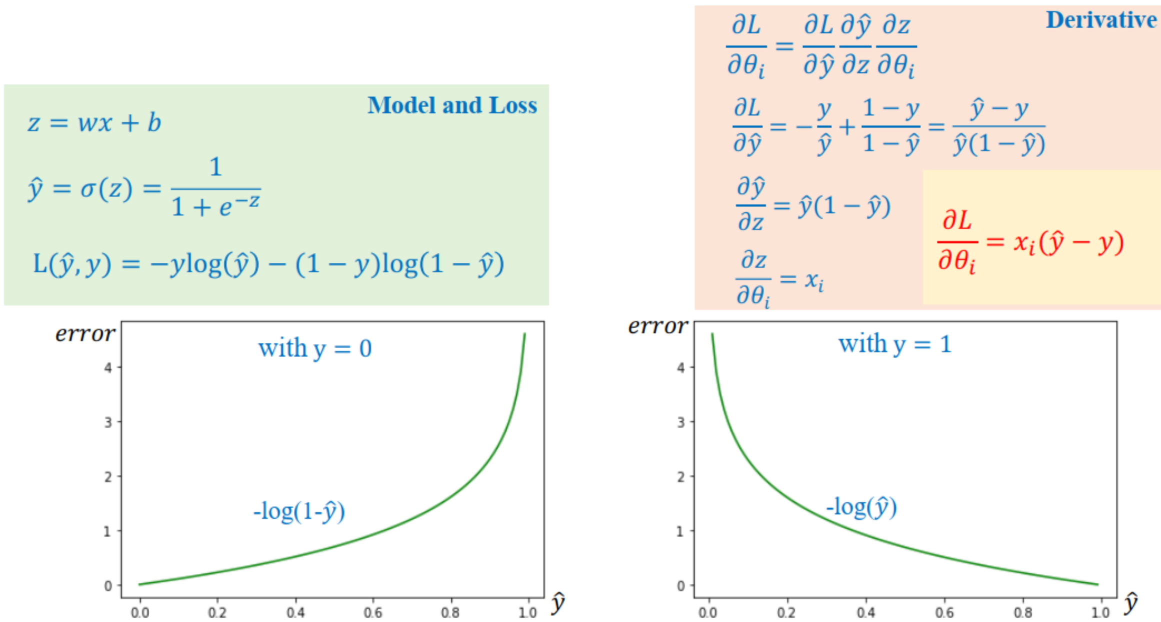
$$\hat{y} = \sigma(z) = \frac{1}{1+e^{-z}}$$

$$\mathrm{L} = (\hat{y} - y)^2$$

$$\frac{\partial L}{\partial \theta_i} = 2x_i(\hat{y} - y)\hat{y}(1 - \hat{y})$$

Từ hình vẽ ta có thể nhìn ra vấn đề phát sinh khi kết hợp MSE với hàm Sigmoid, hàm mất mát tạo ra trở nên **không lồi** (non-convex). Điều này có nghĩa là nó có nhiều điểm cực tiểu cục bộ. Gradient Descent có thể bị "mắc kẹt" tại một điểm cực tiểu cục bộ thay vì tìm ra điểm cực tiểu toàn cục, dẫn đến hiệu suất mô hình không tối ưu.

2.2.2: Binary Cross-Entropy (BCE):



Khi BCE kết hợp cùng với hàm Sigmoid tạo nên hàm mất mát lỗi hoàn hảo(Convex). Và vì hàm lồi chỉ có một điểm cực tiểu toàn cục duy nhất nên đảm bảo rằng thuật toán sẽ luôn hội tụ về giải pháp tối ưu.

Đặc điểm	Hồi quy Logistic với MSE	Hồi quy Logistic với BCE
Công thức Mất mát	$L = (\hat{y} - y)^2$	$L = -y\log(\hat{y}) - (1-y)\log(1 - \hat{y})$
Tính chất	Không lồi (Non-convex)	Lồi (Convex)
Tối ưu hóa	Có thể bị kẹt ở cực tiểu cục bộ	Đảm bảo tìm thấy cực tiểu toàn cục
Đạo hàm (theo θ)	$2x(\hat{y} - y)\hat{y}(1 - \hat{y})$	$x(\hat{y} - y)$

3. Triển khai vectorization cho Logistic Regression
   Trong phần tìm hiểu này, chúng ta cùng so sánh giữa Hồi quy Tuyến tính và Hồi quy Logistic đối với hình thức truyền thống và vectorizatiion,

3.1. Vectorization cho 1 sample

1) Pick a sample (x, y) from training data

2) Compute the output $\hat{y}$

$$z = wx + b \qquad \hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

3) Compute loss

$$L(\hat{y}, y) = (-y\log\hat{y} - (1-y)\log(1-\hat{y}))$$

4) Compute derivative

Traditional

$$\frac{\partial L}{\partial w} = x(\hat{y} - y) \qquad \frac{\partial L}{\partial b} = (\hat{y} - y)$$

5) Update parameters

$$w = w - \eta\frac{\partial L}{\partial w} \qquad b = b - \eta\frac{\partial L}{\partial b}$$

η is learning rate

1) Pick a sample $(\mathbf{x}, y)$ from training data

2) Compute output $\hat{y}$

$$z = \boldsymbol{\theta}^T\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{\theta} \qquad \hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

3) Compute loss

$$L(\hat{y}, y) = (-y\log\hat{y} - (1-y)\log(1-\hat{y}))$$

Vectorized

4) Compute derivative

$$\nabla_{\boldsymbol{\theta}} L = \boldsymbol{x}(\hat{y} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta\nabla_{\boldsymbol{\theta}} L$$

η is learning rate

1) Pick a sample (x, y) from training data

2) Compute the output $\hat{y}$

$$\hat{y} = wx + b$$

3) Compute loss

$$L = (\hat{y} - y)^2$$

Simple

4) Compute derivative

$$\frac{\partial L}{\partial w} = 2x(\hat{y} - y) \qquad \frac{\partial L}{\partial b} = 2(\hat{y} - y)$$

5) Update parameters

$$w = w - \eta\frac{\partial L}{\partial w} \qquad b = b - \eta\frac{\partial L}{\partial b}$$

η is learning rate

1) Pick a sample (x, y) from training data

2) Compute output $\hat{y}$

$$\hat{y} = \boldsymbol{\theta}^T\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{\theta}$$

3) Compute loss

$$L = (\hat{y} - y)^2$$

Vectorized

4) Compute derivative

$$L'_{\boldsymbol{\theta}} = 2\boldsymbol{x}(\hat{y} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta L'_{\boldsymbol{\theta}}$$

η is learning rate

28

3.2. Vectorization for N sample

1) Pick all the samples from training data

2) Compute output $\hat{\boldsymbol{y}}$

$$z = \boldsymbol{x\theta}$$

$$\hat{\boldsymbol{y}} = \sigma(\boldsymbol{z}) = \frac{1}{1 + e^{-\boldsymbol{z}}}$$

3) Compute loss

$$L(\hat{\mathbf{y}}, \mathbf{y}) = \frac{1}{\mathrm{N}}\left(-\mathbf{y}^{\mathrm{T}}\log\hat{\mathbf{y}} - (\mathbf{1}-\mathbf{y})^{\mathrm{T}}\log(\mathbf{1}-\hat{\mathbf{y}})\right)$$

4) Compute derivative

$$\nabla_{\boldsymbol{\theta}} L = \frac{1}{\mathrm{N}}\boldsymbol{x}^{\mathrm{T}}(\hat{\mathbf{y}} - \boldsymbol{y})$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_{\boldsymbol{\theta}} L$$

1) Pick all the N samples from training data

2) Compute output $\hat{\mathbf{y}}$

$$\hat{\boldsymbol{y}} = \mathrm{X}\boldsymbol{\theta}$$

3) Compute loss

$$L = \frac{1}{N}(\hat{\boldsymbol{y}} - \boldsymbol{y})^{\mathrm{T}}(\hat{\boldsymbol{y}} - \boldsymbol{y})$$

4) Compute derivative

$$\boldsymbol{k} = 2(\hat{\boldsymbol{y}} - \boldsymbol{y})$$

$$\nabla_{\boldsymbol{\theta}}\mathrm{L} = \mathrm{X}^{T}\boldsymbol{k}$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta\frac{\nabla_{\boldsymbol{\theta}}\mathrm{L}}{N}$$

η is learning rate

4. Kết luận:

Hồi quy Logistic là một cải tiến quan trọng từ Hồi quy Tuyến tính, cho phép giải quyết các bài toán phân loại nhị phân một cách hiệu quả bằng cách sử dụng **hàm Sigmoid** để giới hạn đầu ra trong khoảng (0, 1). Yếu tố quyết định sự thành công của thuật toán là việc lựa chọn hàm mất mát. **Binary Cross-Entropy (BCE)** vượt trội hoàn toàn so với Sai số Bình phương Trung bình (MSE), vì nó đảm bảo một bài toán tối ưu hóa **lồi**, giúp thuật toán hội tụ một cách đáng tin cậy.