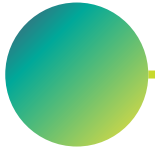




VietAI

Softmax Regression Model Evaluation



VietAI Teaching Team



VietAI

Softmax Regression

Softmax Regression

1. Nhắc lại Logistic Regression
2. Softmax Regression Model
3. Cost Function
4. Một số tính chất của Softmax Regression

1 Nhắc lại Logistic Regression

- Giải bài toán Classification, đơn giản nhất là Binary Classification.
- Ví dụ: Phân loại bài thi cuối khoá của sinh viên đậu hay rớt dựa trên số giờ học và số đề thi mẫu sinh viên đã làm trước khi thi.

Số giờ ôn tập	Số đề mẫu đã giải	Đậu
0.50	0	0
1.00	1	0
1.75	4	1
1.75	3	0
2.00	6	1
3.25	2	0
3.50	5	1
4.25	4	1

1 Nhắc lại Logistic Regression

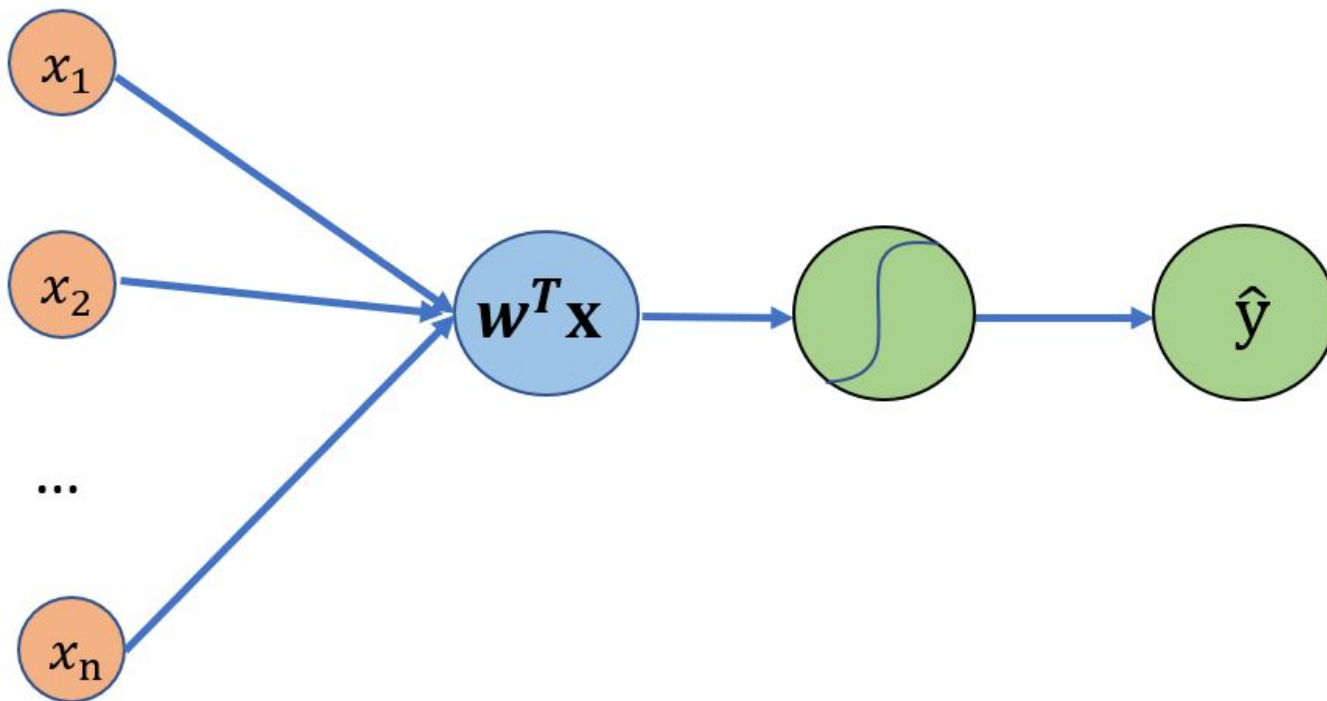
- Logistic Regression giải bài toán Classification bằng cách đưa về bài toán dự đoán xác suất của lớp positive:

$$P(y = 1|x) = h_w(x) = g(w^\top x) = \frac{1}{1 + \exp(-w^\top x)}$$

- Cost Function của Logistic Regression:

$$J(w) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h_w(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_w(x^{(i)}))]$$

1 Nhắc lại Logistic Regression

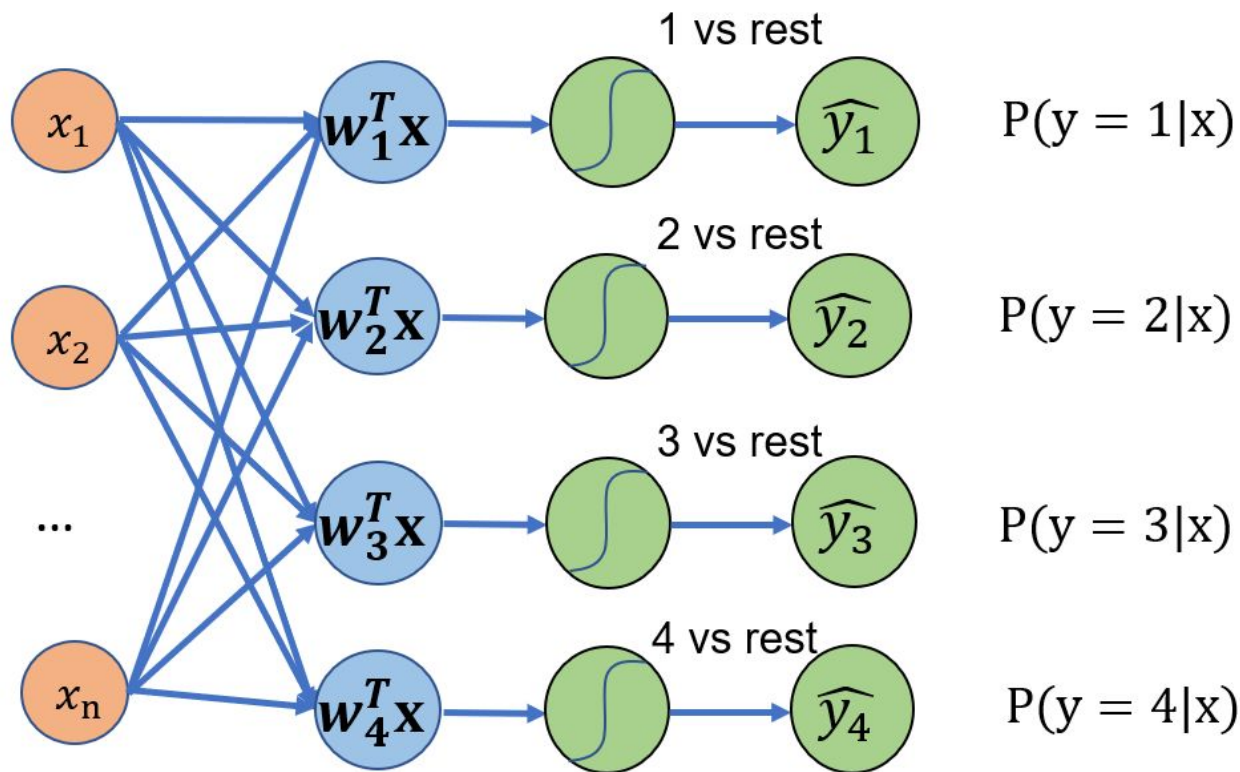


2 Multiclass Classification - Introduction

- Ví dụ: Phân loại bài thi cuối khoá của sinh viên thuộc loại giỏi, khá, trung bình hay yếu dựa trên số giờ học và số đề mẫu sinh viên đã làm trước khi thi.

Số giờ ôn tập	Số đề mẫu đã giải	Xếp loại bài thi	Lớp
0.50	0	Yếu	1
1.00	1	Trung bình	2
1.75	4	Trung bình	2
1.75	3	Khá	3
2.00	6	Giỏi	4
3.25	2	Khá	3
3.50	5	Giỏi	4
4.25	4	Giỏi	4

2 Multiclass Classification - Log. Reg.



2 Softmax Regression Model

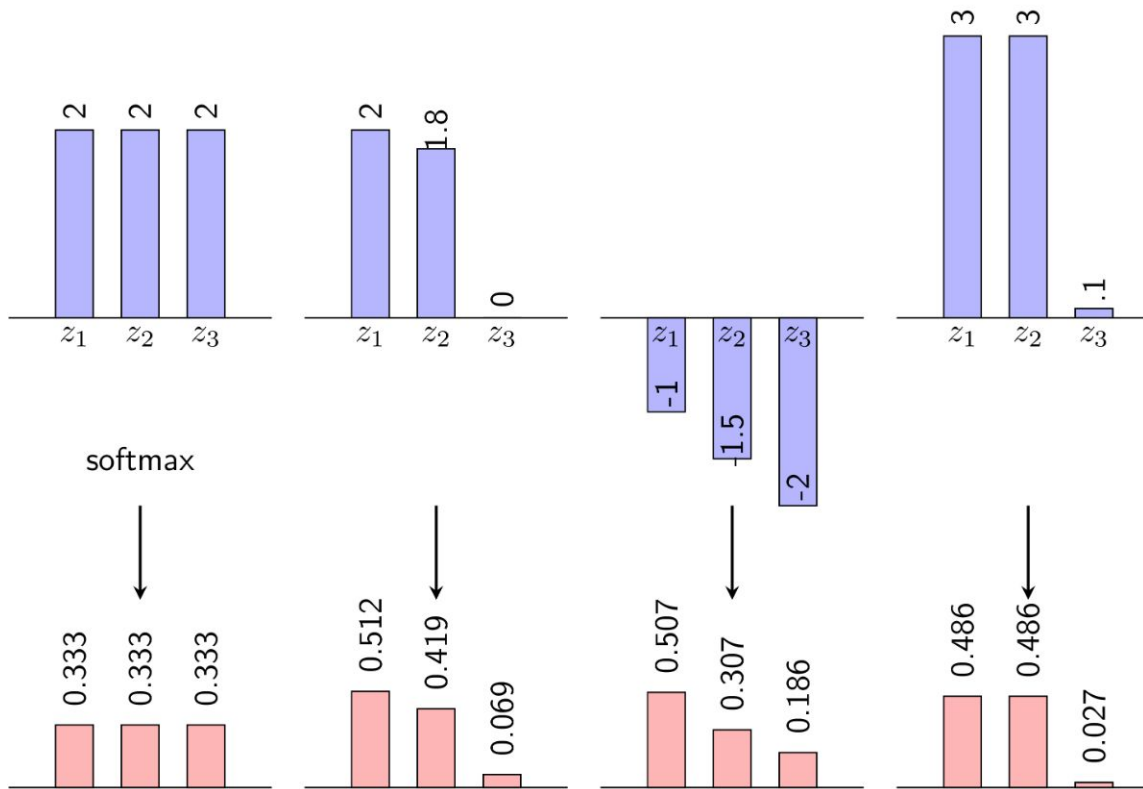
- Với bài toán Multiclass Classification, khi đó y có thể nhận K giá trị khác nhau (tương ứng với K lớp) $y^{(i)} \in \{1, 2, \dots, K\}$. Mô hình Softmax Regression dự đoán xác suất của từng lớp k khi biết input vector x :

$$P\left(y^{(i)} = k | x^{(i)}; W\right) = \frac{\exp\left(w^{(k)\top} x^{(i)}\right)}{\sum_{j=1}^K \exp\left(w^{(j)\top} x^{(i)}\right)}$$

- Phần tử $\sum_{j=1}^K \exp\left(w^{(j)\top} x^{(i)}\right)$ được dùng để **chuẩn hóa (normalize)** phân phối xác suất để tổng xác suất của các lớp bằng 1: $\sum_{k=1}^K P(y^{(i)} = k | x^{(i)}; W) = 1$

2

Softmax Regression - Input and Output



Nguồn: Machine Learning cơ bản
VietAI ML Foundation Class 6 - Lecture 6

2 Softmax Regression - Weights

- Như vậy mô hình Softmax Regression có k bộ trọng số $w^{(1)}, w^{(2)}, \dots, w^{(K)} \in \mathbb{R}^n$
- Để dễ tính toán và biểu diễn, người ta biểu diễn các bộ trọng số trên của mô hình dưới dạng ma trận $W \in \mathbb{R}^{n \times K}$

$$W = \begin{bmatrix} | & | & & | \\ w^{(1)} & w^{(2)} & \dots & w^{(K)} \\ | & | & & | \end{bmatrix}$$

- Khi đó mô hình có thể biểu diễn dưới dạng tích giữa ma trận W và input vector x .

3 Softmax Regression - Cost Function

- Cost Function của Softmax Regression cũng được xây dựng dựa trên việc cực đại tích xác suất (mô hình dự đoán) của các phân lớp đúng.

$$1\{A\} = \begin{cases} 0 & A \text{ is False} \\ 1 & A \text{ is True} \end{cases} \quad J(W) = -\frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K 1\{y^{(i)} = k\} \log P\left(y^{(i)} = k | x^{(i)}; W\right)$$

$$= -\frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K 1\{y^{(i)} = k\} \log \frac{\exp(w^{(k)\top} x^{(i)})}{\sum_{j=1}^K \exp(w^{(j)\top} x^{(i)})}$$

- Đạo hàm của cost function trong Softmax Regression:

$$\nabla_{w^{(k)}} J(W) = -\frac{1}{m} \sum_{i=1}^m \left(1\{y^{(i)} = k\} - P\left(y^{(i)} = k | x^{(i)}; W\right) \right) x^{(i)}$$

4 Một số tính chất (đọc thêm)

- Nếu trừ tất cả các vector $w^{(1)}, w^{(2)}, \dots, w^{(K)}$ cho vector ψ bất kì ta thu được bộ trọng số mới tương đương với bộ trọng số cũ.

$$\begin{aligned} P\left(y^{(i)} = k | x^{(i)}; W\right) &= \frac{\exp\left(\left(w^{(k)} - \psi\right)^\top x^{(i)}\right)}{\sum_{j=1}^K \exp\left(\left(w^{(j)} - \psi\right)^\top x^{(i)}\right)} \\ &= \frac{\exp\left(w^{(k)\top} x^{(i)}\right) \exp\left(-\psi^\top x^{(i)}\right)}{\sum_{j=1}^K \exp\left(w^{(j)\top} x^{(i)}\right) \exp\left(-\psi^\top x^{(i)}\right)} \\ &= \frac{\exp\left(w^{(k)\top} x^{(i)}\right)}{\sum_{j=1}^K \exp\left(w^{(j)\top} x^{(i)}\right)} \end{aligned}$$

4 Một số tính chất (đọc thêm)

- Cost function của Logistic Regression có thể được viết lại như sau:

$$\begin{aligned} J(w) &= -\frac{1}{m} \sum_{i=1}^m \left((1 - y^{(i)}) \log (1 - h_w(x^{(i)})) + y^{(i)} \log h_w(x^{(i)}) \right) \\ &= -\frac{1}{m} \sum_{i=1}^m \sum_{k=0}^1 1 \{y^{(i)} = k\} \log P(y^{(i)} = k | x^{(i)}; w) \end{aligned}$$

- Với trường hợp $K = 2$, Softmax Regression có thể được viết như sau:

$$\begin{bmatrix} P(y = 1|x) \\ P(y = 2|x) \end{bmatrix} = \frac{1}{\exp(w^{(1)\top}x) + \exp(w^{(2)\top}x)} \begin{bmatrix} \exp(w^{(1)\top}x) \\ \exp(w^{(2)\top}x) \end{bmatrix}$$

4 Một số tính chất (đọc thêm)

- Như vậy có thể thấy, khi đó Softmax Regression chính là Logistic Regression với $w' = w^{(2)} - w^{(1)}$

$$\begin{aligned}
 \left[\frac{P(y=1|x)}{P(y=2|x)} \right] &= \frac{1}{\exp((w^{(1)} - w^{(2)})^\top x^{(i)}) + \exp(\vec{0}^\top x)} [\exp((w^{(1)} - w^{(2)})^\top x) \exp(\vec{0}^\top x)] \\
 &= \left[\frac{\frac{1}{1 + \exp((w^{(1)} - w^{(2)})^\top x^{(i)})}}{\frac{\exp((w^{(1)} - w^{(2)})^\top x)}{1 + \exp((w^{(1)} - w^{(2)})^\top x^{(i)})}} \right] \\
 &= \left[1 - \frac{1}{1 + \exp((w^{(1)} - w^{(2)})^\top x^{(i)})} \right]
 \end{aligned}$$

Model Evaluation

Model Evaluation

1. Testing
2. Cross Validation - Chọn lựa mô hình và hyperparameters
3. Cách giải quyết một số vấn đề trong Machine Learning

1 Testing

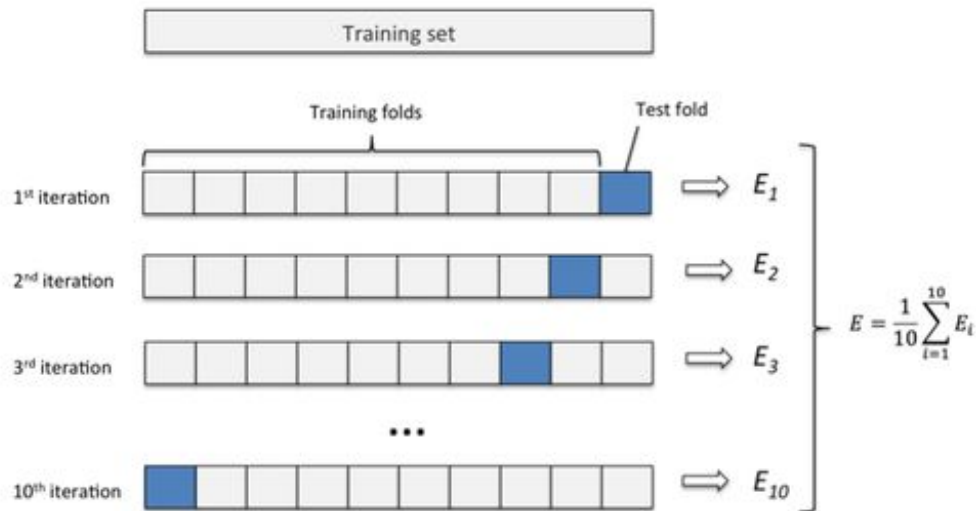
- Testing giúp đánh giá khả năng hoạt động của mô hình khi đầu vào là dữ liệu không xuất hiện trong tập training.
- Sử dụng tập test để tính độ lỗi, accuracy, F1-Score... và so sánh với giá trị tương ứng trên training set.
- Thông thường tập dữ liệu được chia thành hai tập: training set (chiếm 80%) và test set (chiếm 20%).

2 Cross Validation

- **Validation** giúp ta kiểm soát được quá trình training và lựa chọn được mô hình, các hyperparameters tốt nhất mà không gặp phải vấn đề overfitting.
- Validation có thể thực hiện bằng cách tạo một **tập validation (validation set)** từ tập dữ liệu ban đầu (chiếm 20%). Thực hiện tính cost function (accuracy, F1-Score ...) **trên tập validation** sau khi đã huấn luyện mô hình với tập training, sau đó chọn mô hình (features, hyperparameters ...) cho kết quả tốt nhất.

2 Cross Validation

- ***K-Fold Cross Validation*** được thực hiện bằng cách chia tập dữ liệu thành K phần bằng nhau.
- Thực hiện K lần việc huấn luyện mô hình, mỗi lần sử dụng $K - 1$ phần để huấn luyện và 1 phần để validation. Kết quả validation được tính bằng cách chia trung bình kết quả của K lần huấn luyện.



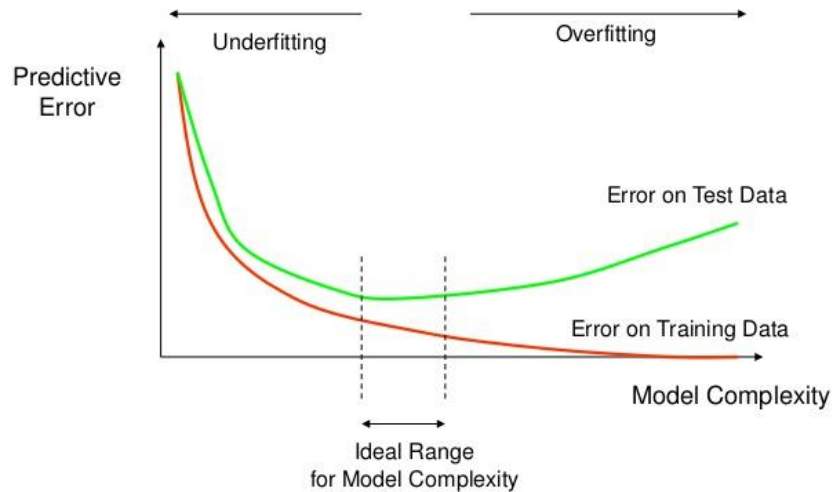
3 Cách giải quyết một số vấn đề

- **Overfitting:**

- Tìm thêm dữ liệu huấn luyện
- Giảm bớt số features
- Tăng regularization parameter (λ)

- **Underfitting:**

- Tăng số features
- Sử dụng mô hình phức tạp hơn
- Giảm regularization parameter (λ)



Tài liệu tham khảo

1. [UFLDL Tutorial - Softmax Regression - Stanford University](#)
2. *Advice for Applying Machine Learning* - Machine Learning course - Andrew Ng
- Coursera
3. Softmax Regression - Tiệp Vũ - Machine Learning cơ bản