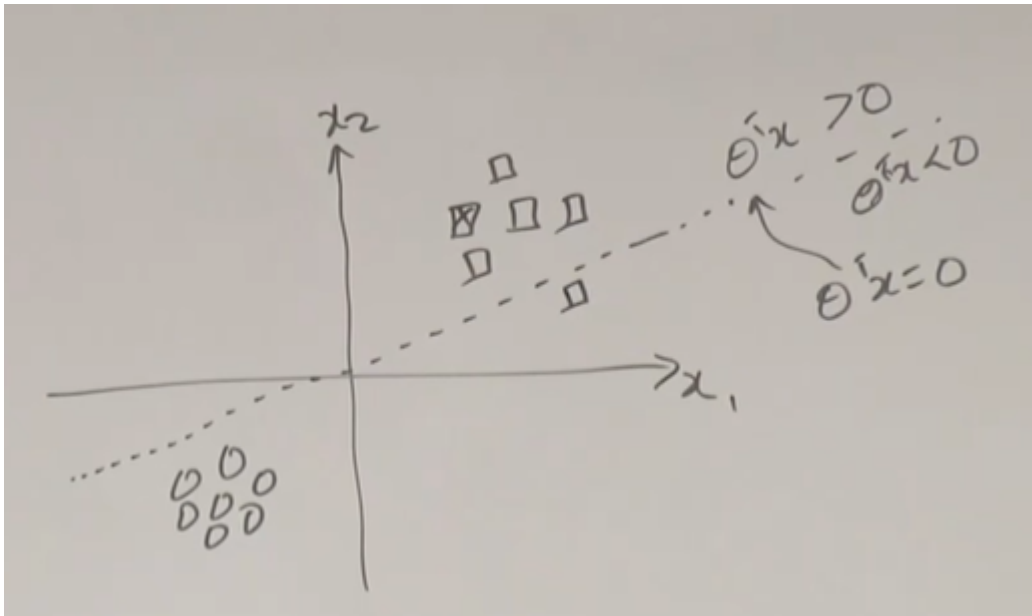


Lecture 4: Perceptron & Generalized Linear Model

Perceptron

Perceptron là một loại mô hình tuyến tính, nghĩa là nó tạo ra dự đoán dựa trên tổng trọng số của các đặc trưng đầu vào. Mô hình này sử dụng một hàm chuyển đổi, thường được gọi là hàm kích hoạt, để quyết định đầu ra. Trong trường hợp của perceptron, hàm kích hoạt là một hàm nhị phân, có nghĩa là nó chỉ có thể trả về một trong hai giá trị.

Perceptron nhận đầu vào là một vector các thuộc tính, ví dụ như đặc điểm của một bức ảnh, một câu, hoặc một dẫn xuất tài chính. Nó sau đó tiến hành nhân mỗi thuộc tính với trọng số tương ứng của nó, sau đó tổng hợp tất cả các kết quả này. Kết quả cuối cùng sau đó được chuyển qua hàm kích hoạt để quyết định đầu ra cuối cùng của mô hình. Đầu ra này có thể được sử dụng để phân loại đầu vào vào một trong hai lớp.



Exponential Family

PDF, hoặc Phân phối xác suất (Probability Density Function), là một hàm mô tả xác suất mà một biến ngẫu nhiên liên tục nằm trong một khoảng nhất định.

Hàm này có giá trị không âm và diện tích dưới đồ thị của nó (tức là tích phân

của nó trên toàn bộ không gian mẫu) bằng 1. Với PDF, chúng ta không thể lấy giá trị xác suất tại một giá trị cụ thể, mà phải tính xác suất trong một khoảng.

Công thức tổng quát của một PDF (hàm mật độ xác suất) được biểu diễn như sau:

$$p(y; \eta) = b(y) \exp(\eta^T T(y) - a(\eta))$$

trong đó:

- **$b(y)$** : Đây là hàm cơ sở (base measure) của y . Hàm này không phụ thuộc vào tham số η và thường được dùng để đảm bảo hàm mật độ xác suất được chuẩn hóa.
- **\exp** : Hàm mũ, một thành phần chính trong công thức của phân phối thuộc họ phân phối mũ.
- **η^T** : Đây là tham số tự nhiên (canonical parameter). Ký hiệu η^T biểu diễn chuyển vị của vector η .
- **$T(y)$** : Đây là hàm thống kê kê đủ (sufficient statistic) của y . Hàm này phụ thuộc vào y và thường xuất hiện trong phân phối thuộc họ phân phối mũ.
- **$a(\eta)$** : Đây là hàm log-normalizer (hoặc log-partition function), đảm bảo rằng hàm mật độ xác suất được chuẩn hóa. Nó phụ thuộc vào η và đảm bảo tổng xác suất bằng 1.

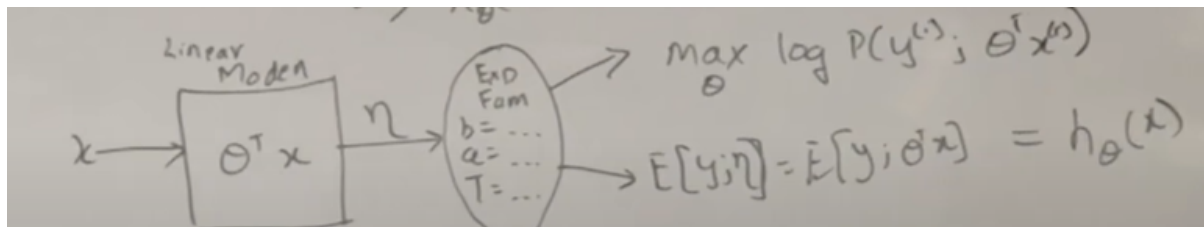
Một số phân phối thuộc Exponential Family bao gồm:

1. Phân phối chuẩn (Normal distribution)
2. Phân phối Poisson (Poisson distribution)
3. Phân phối nhị thức (Binomial distribution)
4. Phân phối gamma (Gamma distribution):
5. Phân phối nhị thức âm (Negative binomial distribution)
6. Phân phối đa thức (Multinomial distribution)
7. Phân phối Beta (Beta distribution)
8. Phân phối Dirichlet (Dirichlet distribution)
9. Phân phối Bernoulli (Bernoulli distribution)

Generalized Linear Model (GLM)

Một số giả định:

- Với $y|x; \theta \sim \text{Exponential Family}(\eta)$
- $\eta = \theta^T x$
-



Generalized Linear Models (GLMs) là một nhóm các mô hình thống kê tổng quát hóa phương trình hồi quy tuyến tính để áp dụng cho nhiều loại dữ liệu khác nhau. GLM mở rộng mô hình hồi quy tuyến tính bằng cách cho phép các biến phụ thuộc có phân phối khác không phải là phân phối chuẩn (Gaussian distribution) và liên kết các biến độc lập với biến phụ thuộc thông qua một hàm liên kết (link function).

GLM bao gồm ba thành phần chính:

1. Hàm phân phối (Distribution Function):

- GLM cho phép biến phụ thuộc Y tuân theo các phân phối thuộc họ exponential (exponential family of distributions), chẳng hạn như phân phối chuẩn, phân phối Poisson, phân phối nhị thức (binomial), phân phối gamma, v.v.

2. Hàm liên kết (Link Function):

- Hàm liên kết xác định mối quan hệ giữa kỳ vọng của biến phụ thuộc $E(Y)=\mu$ và tổ hợp tuyến tính của các biến độc lập $\eta=X\beta$. Một số hàm liên kết phổ biến bao gồm:

- Hàm đồng nhất (Identity link): $g(\mu) = \mu$
- Hàm log (Log link): $g(\mu) = \log(\mu)$
- Hàm logit (Logit link):

$$g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$$

3. Tổ hợp tuyến tính (Linear Predictor):

- Tổ hợp tuyến tính của các biến độc lập được biểu diễn bởi $\eta = X\beta$, trong đó X là ma trận các biến độc lập, β là vector các hệ số hồi quy.

Softmax Regression (Cross entropy)

Có một số định nghĩa sau:

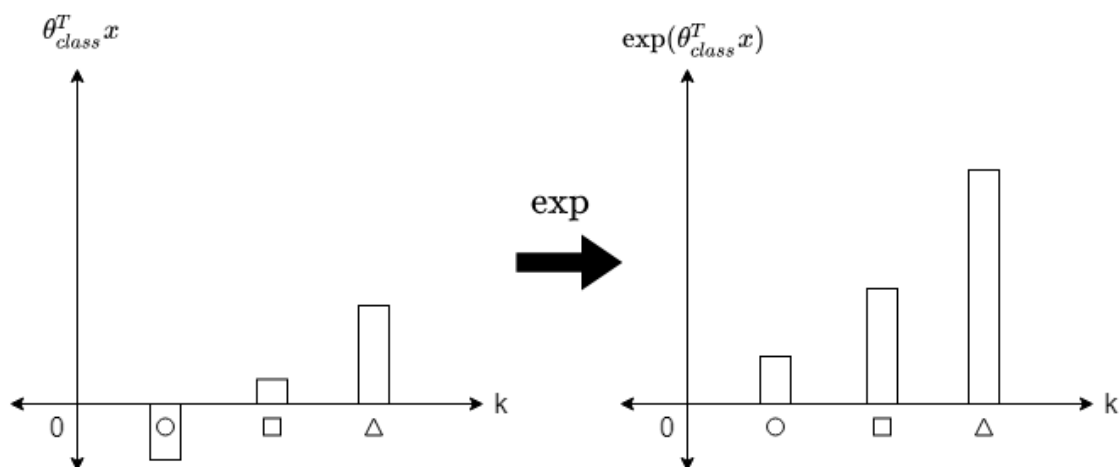
- K - số class của mô hình
- $x^{(i)} \in R^n$
- Nhãn $y = [\{0, 1\}^k]$ hay $y = [0, 0, \dots, 1, 0]$

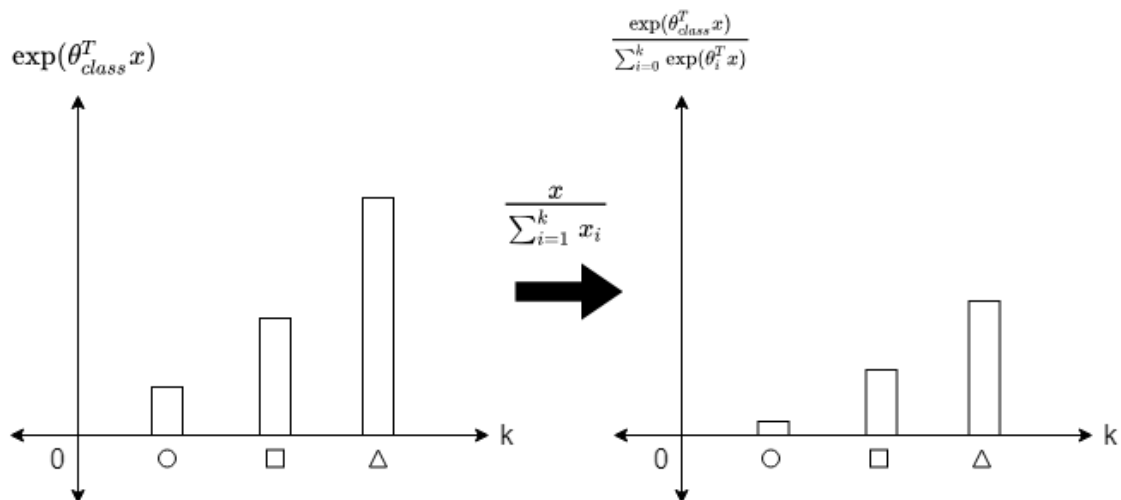
Mỗi nhãn sẽ có một tập các tham số θ khác nhau

Giả sử tập tham số θ sẽ cho ta một giá trị output nằm trong khoảng $[-\infty, +\infty]$, chúng ta có thể chuẩn hóa các giá trị này bằng công thức

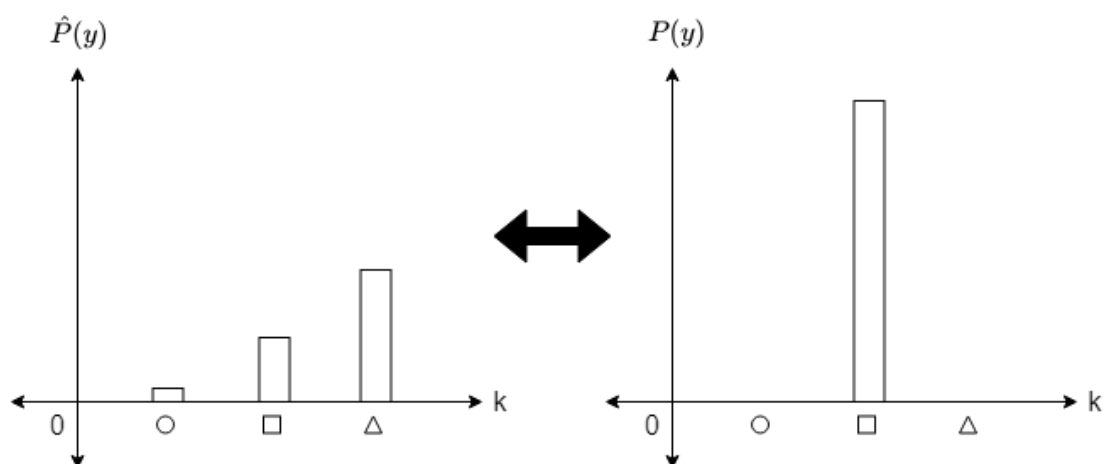
$$\frac{e^{\theta^T x}}{\sum_{j=1}^K e^{\theta^T x}}$$

Công thức trên còn được gọi là công thức softmax. Công thức này sẽ chuẩn hóa các giá trị như hình sau





Đối với mỗi điểm dữ liệu, chúng ta có phân phối xác suất của nhãn giữa mô hình và nhãn thực như sau:



Với công thức softmax chúng ta muốn minimize một giá trị còn được gọi là giá trị cross entropy. Thông thường, giá trị entropy giữa một

$$L(y, \hat{y}) = - \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(\hat{y}_{ij})$$

Trong đó:

- N là số lượng mẫu.
- C là số lớp.

- y_{ij} là nhãn thực tế của mẫu thứ i cho lớp j (có giá trị 0 hoặc 1, với đúng một giá trị 1 và các giá trị còn lại là 0 cho mỗi mẫu).
- \hat{y}_{ij} là xác suất dự đoán của mô hình đối với mẫu thứ i cho lớp j (có giá trị trong khoảng từ 0 đến 1, và tổng các giá trị \hat{y}_{ij} cho mỗi mẫu i bằng 1).