

MaSSP Lecture 2 (part 2)

Kieu Anh Le

July 2019

1 Review toán nền tảng:

1.1 Linear function

Hàm tuyến tính $F: X \xrightarrow{f} Y$
hay $Y = f(X)$

- Trường hợp 1: $X \in R^d; Y \in R$
 $\Rightarrow Y = a^T \cdot x + b (a \in R^d, b \in R)$
 $y = a^T \cdot x + b = \dot{a}^T \cdot \dot{X}$
Trong đó: $\dot{a} = \begin{bmatrix} b \\ a \end{bmatrix}; \dot{X} = \begin{bmatrix} 1 \\ X \end{bmatrix}$
- Trường hợp 2: $X \in R^d; Y \in R^k$
 $\Rightarrow Y = W_x + b (W \in R^{k \times d})$ W là 1 ma trận
Có thể viết: $Y = \dot{W} \cdot \dot{X}$ Trong đó: $\dot{W} = \begin{bmatrix} w_1 \\ \dots \\ w_{d+1} \end{bmatrix}; \dot{X} = \begin{bmatrix} 1 \\ \dots \\ X \end{bmatrix}$

1.2 Non-linear function

- Để cố định 1 khoảng giá trị cho output -> **Dùng hàm sigmoid**
VD: $y \in R \rightarrow \sigma(y) \in (0, 1)$
 $\Rightarrow 2\sigma(y) - 1 \in (-1, 1)$
- Để loại bỏ những giá trị không mong muốn (VD: loại bỏ giá trị < 0 , chỉ lấy phần dương)
-> Dùng hàm ReLU: $\max(0, 1)$
- Softmax Function: Là hàm để tính xác suất của nhiều giá trị

2 Softmax Regression:

2.1 Softmax Function:

Từ hàm sigmoid: $\sigma(y) = \sigma \begin{bmatrix} y_1 \\ \dots \\ y_k \end{bmatrix} = \begin{bmatrix} \sigma(y_1) \\ \dots \\ \sigma(y_k) \end{bmatrix}$

=> Normalize: $\frac{\sigma(y_1)}{\sum_{i=1}^k \sigma(y_i)}$

Với $p = \text{softmax}(y) =>$

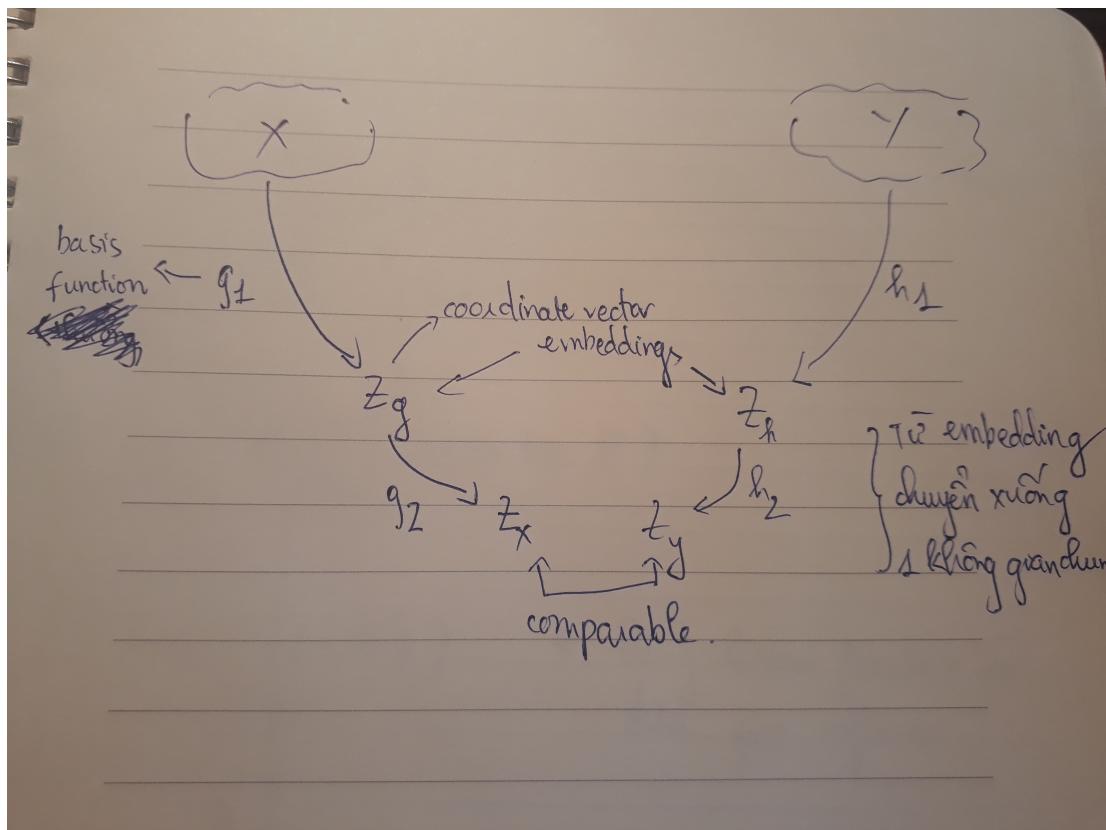
$0 \leq p_i \leq 1$

$\sum_{i=1}^k p_i = 1$

=> Tác dụng: Sử dụng để chuyển đổi từ vector tọa độ thành vector xác suất

2.2 Bài toán Softmax Regression:

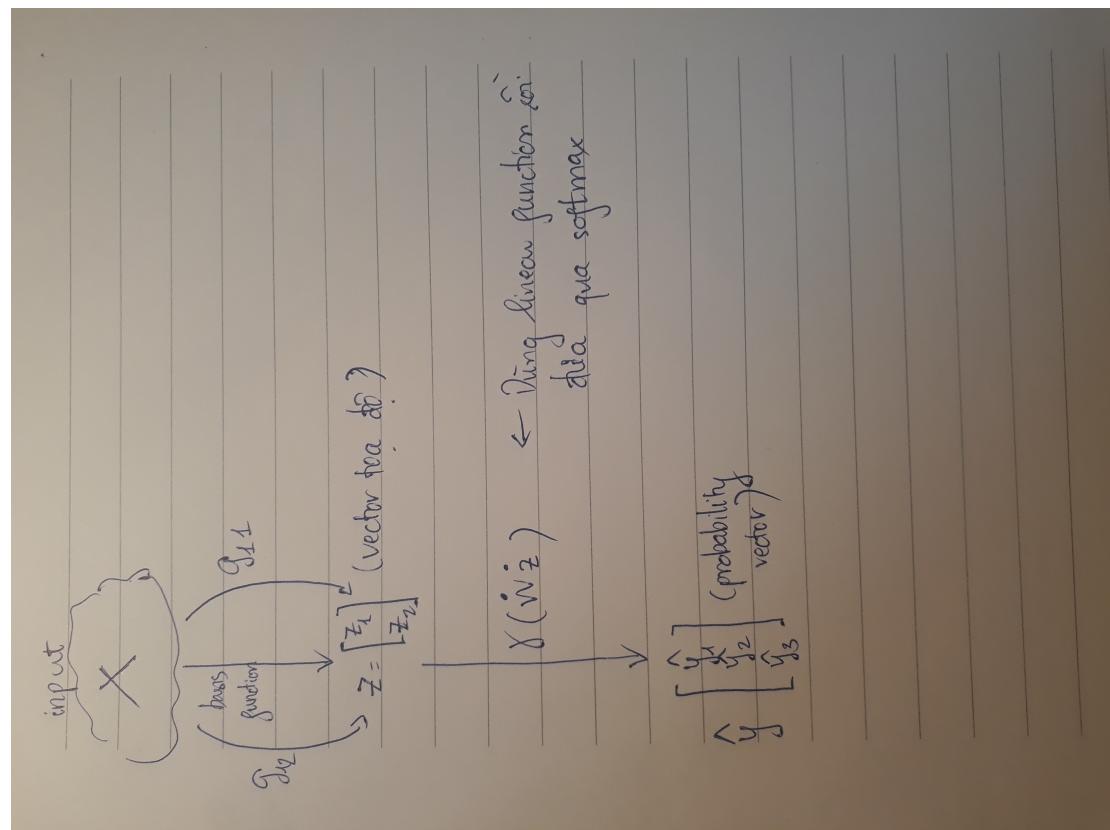
2.2.1 Unified Framework:



Hình 1: Unified Framework

- Basis function g1: dùng linear function
- g2: dùng non-linear function (sigmoid, ReLU, ...) -> Có tác dụng trích xuất ra Z_x có thể so sánh được với Z_y

2.2.2 Giải thích bài toán Softmax Regression:

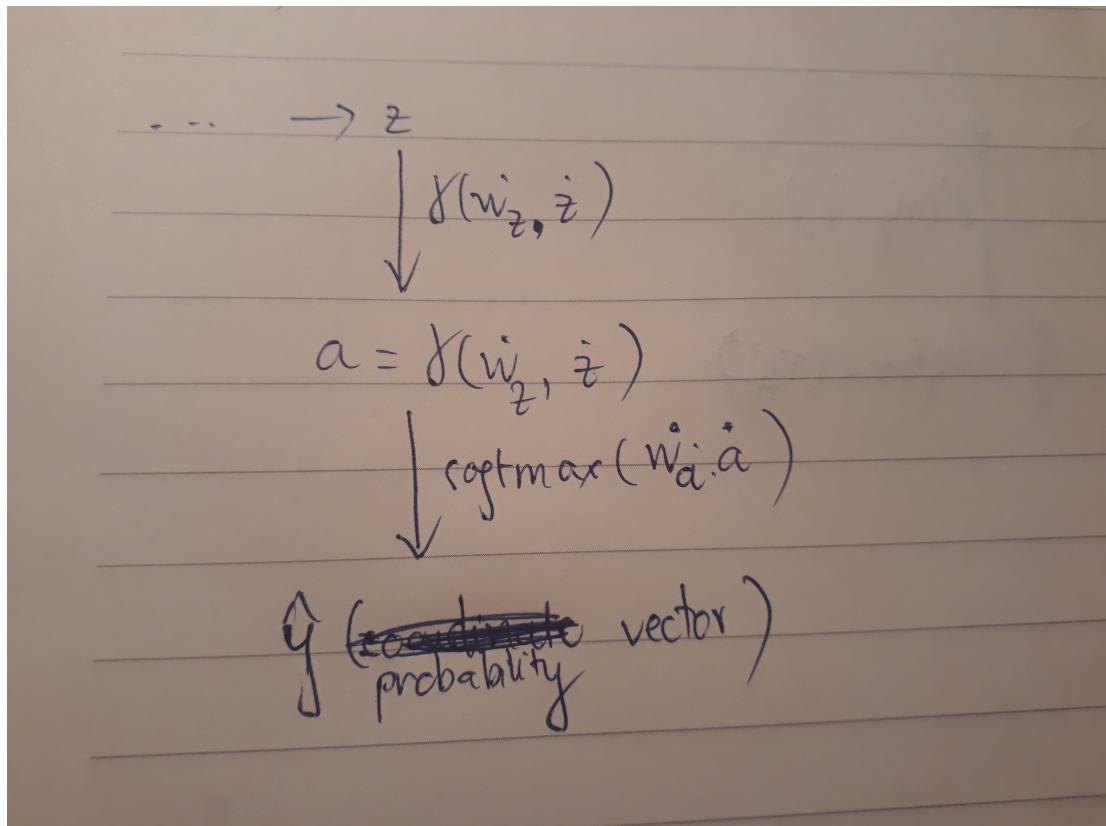


Hình 2: Unified Framework - Softmax Regression

- Mục tiêu: Vẽ đường chia cắt (decision boundary) những phần dữ liệu thuộc 3 class khác nhau => Tie-breaking: Tìm tập hợp những điểm mà probability của 2 class bằng nhau (không tính được argmax của \hat{y}_i)
- Cách tính xác suất lớn nhất: argmax \hat{y}_i , $i=1 \rightarrow k$

3 Multi-layer Perceptron

Bài toán: Vẽ những đường cong để phân chia các class dữ liệu không phân chia
được bằng đường thẳng (linearly non-separable)
=> Cần dùng non-linear function



Hình 3: Unified Framework - Multi-layer Regression

- Từ z , tính a theo non-linear function -> Từ a tính probability vector \hat{y} theo non-linear function tiếp theo (ở đây là softmax)