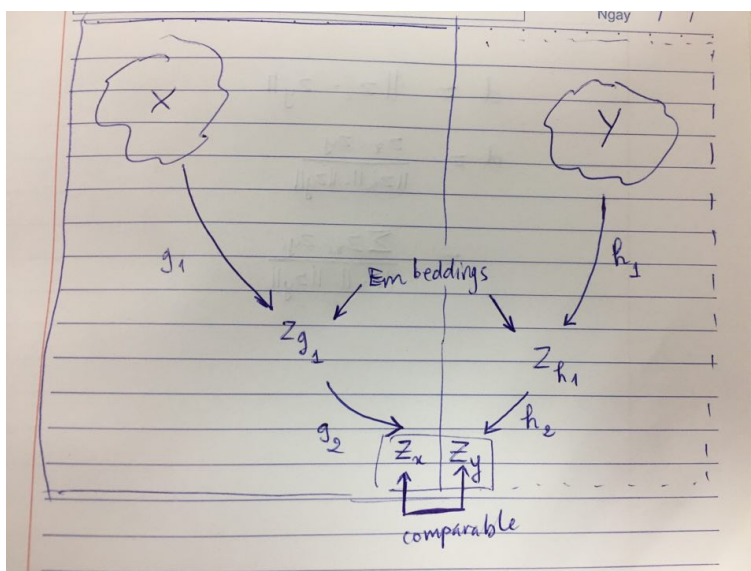


# MASSP

Nguyễn Thiện Nhân  
Phổ Thông Năng Khiếu - T1619

2 July 2019

## 1 Mô hình chung



Hình 1: Mô hình chung machine learning

Mô hình chung là tìm hàm  $f : X \rightarrow Y$  với  $X$  là bộ cơ sở dữ liệu và  $Y$  là các đầu ra sao cho hàm  $f$  với bộ dữ liệu  $X$  thì cho đầu ra  $Y$  hiệu quả nhất.

Một mô hình chung có thể biểu diễn như sau :

$$X \xrightarrow{h_1} Z_{h_1} \xrightarrow{h_2} Z_X \textbf{(1)}$$

$$Y \xrightarrow{g_1} Z_{g_1} \xrightarrow{g_2} Z_Y \textbf{(2)}$$

Với  $h_1, g_1, h_2, g_2$  là các **basis function**.

Hàm  $h_2, g_2$  dùng để đưa các  $Z_{g_1}, Z_{h_1}$  và tạo ra được tương ứng  $Z_X, Z_Y$  vào không gian hàm chung. Từ đó đưa ra so sánh giữa  $Z_X, Z_Y$ . Để so sánh ta dùng

$$d(Z_X, Z_Y) = P(X, Y_{g_1, g_2, h_1, h_2})$$

Trong đó  $P$  là hàm để đo **Performance Measure**  
 Có nhiều cách tính  $d$  và  $P$ , chẳng hạn

$$d(X_1, X_2) = \sqrt{\sum (x_{ij_1} - x_{ij_2})^2} \quad (3)$$

$$d(X_1, X_2) = \frac{\sum x_{ij_1} \cdot x_{ij_2}}{||x_{ij_1}|| \cdot ||x_{ij_2}||} \quad (4)$$

$$d(X_1, X_2) = -\sum p_{X_1 i} \log(p_{X_2 i})$$

## 2 Linear Regression

Ta sẽ lấy ví dụ theo mô hình trên.

Lấy  $X$  là tập hợp các tính chất của một căn nhà ví dụ vị trí địa lý, kích thước, độ bền,....

$Y$  là tập hợp giá trị của căn nhà.

Việc sử dụng Linear Regression theo mô hình trên là việc mã hoá  $Z_{g_1}$ . Như mô tả ở mô hình chung, khi đưa  $X$  qua basis function  $h_1$  sẽ cho ta được các toạ độ (coordinates)  $Z_{g_1}$ . Từ các toạ độ, bước  $h_2$  sẽ là từ  $Z$  ta nhân vô hướng với một vector  $w$  cột ví dụ như

$$w = [r_1, r_2, \dots, r_N]^T$$

để dự đoán giá nhà  $\hat{y}$  và so sánh nó với giá trị thực (linear). Ở dạng công thức toán ta có thể viết là

$$\hat{y} = Z^T w$$

Từ đây ta điều chỉnh  $d(y, \hat{y})$  (**Performance Measure**) sao cho sai số của đầu ra đạt giá trị nhỏ nhất. Hay nói cách khác là ta cần tìm  $w^*$  sao cho

$$w^* = \underset{i=1}{\operatorname{argmin}} \sum^N d(y, \hat{y})$$

Với tập dữ liệu  $D = \{(x_i, y_i)\}_{i=1}^N$

## 3 Classification

Một số hàm toán sử dụng

Hàm **sigmoid**

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

Hàm **softmax**

$$a_i = \frac{\sigma(x_i)}{\sum_{i=1}^n \sigma(x_i)}$$

Hàm **ReLU**

$$f(g(x)) = \begin{cases} 0, & g(x) < 0 \\ g(x), & g(x) \geq 0 \end{cases}$$

### 3.1 Linear Classification

Đối với Linear Classification, ta có 2 thuật toán cơ bản của bài toán là Logistic Regression và Softmax Regression.

Bài toán Logistic Regression thường sử dụng để giải quyết các bài toán **binary classification**. Còn bài toán Softmax Regression thì sử dụng để giải quyết những bài toán có nhiều class hơn.

Theo mô hình chung trên, nguyên tắc hoạt động chung của hai hàm là ta được cung cấp dữ liệu là vector tọa độ  $Z_{g_1}$  và ta cần đưa đầu ra là  $Z_Y$ . Để làm được như vậy trước hết ta đưa  $Z_{g_1}$  qua một **basis function**  $g_2$  (linear) và sau đó điều chỉnh  $d(y, \hat{y})$  để cho ra được một vector  $\hat{y}$  là các **probability vector**.

**Định nghĩa : probability vector** là vector  $\hat{y} = [p_{y_1}, p_{y_2}, \dots, p_{y_n}]^T$  (với Linear Classification) thỏa các điều kiện sau

- $(0 \leq p_i \leq 1)_{i=1}^n$
- $\sum_{i=1}^n p_i = 1$

Để so sánh  $\hat{y}$  và  $y$  ta cũng cần chuyển vector  $y$  về probability vector dùng **hot coding**. Hàm để tính  $d(y, \hat{y})$  (Performance Measure) là hàm **cross-entropy** với công thức

$$d(y, \hat{y}) = - \sum_{i=1}^n y_i \log(\hat{y}_i)$$

### 3.2 Nonlinear Classification (Multi-layer Perceptron)

Xét về mô hình chung thì Nonlinear và Linear cũng có nhiều điểm tương đồng như việc tính  $d(y, \hat{y})$  hay đầu ra  $\hat{y}$  đều là probability vector Xét theo mô hình chung của machine learning, điểm khác nhau giữa Nonlinear và Linear là ở bước  $g_2$ . Bước  $g_2$  của Nonlinear sẽ được thông qua thêm nhiều bước ẩn (hidden layers). Biểu diễn của toán học là

$$\begin{aligned} a &= \gamma(\dot{W}_z \dot{z}) \\ \hat{y} &= (\dot{W}_a \dot{a}) \end{aligned}$$

Trong đó hàm  $\gamma(x)$  là một hàm phi tuyến (nonlinear)