

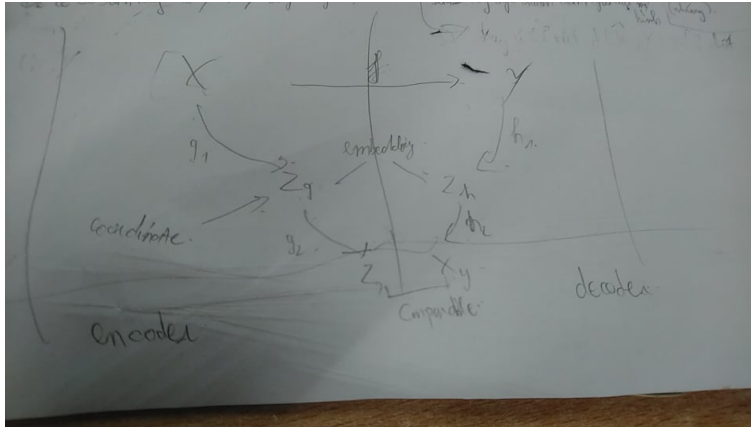
Ngày 1

Huỳnh Vũ Khôi Nguyễn

## 1 Mô hình chung của machine learning

Mô hình chung của machine learning là tìm các hàm  $f : X \rightarrow Y$  với  $X$  là tập dữ liệu đầu vào và  $Y$  là các đặc trưng ta rút ra được từ data. Đây là một mô hình được đơn giản hóa nhưng thể hiện được tính thống nhất cho các mô hình machine learning (simplified but unified)

Mục tiêu của ta là tìm hàm  $f$  hợp lý nhất để mô tả sự liên hệ giữa  $X$  và  $Y$  hay  $f(X) = \hat{Y} \approx Y$  hay  $d(\hat{Y}, Y)$  nhỏ nhất với  $d$  là một chuẩn đánh giá nào đó.



Hình 1: Mô hình chung của machine learning

Giả sử ta có các hàm cơ sở  $(f_1, f_2, \dots, f_n)$  với mỗi hàm thể hiện trích xuất từ  $X$  thì khi đó ta có thể viết  $X$  dưới dạng  $X = y_1 f_1 + y_2 f_2 + \dots + y_n f_n$ . Khi đó  $(y_1, y_2, \dots, y_n)$  được gọi là coordinate.

Kết quả của các trích xuất từ  $X$  có thể không cùng một chuẩn với cái ta quan sát được ở  $Y$  nên ta cần chuẩn hóa tức là đưa chúng về cùng một dạng mà ta có thể so sánh được. Quá trình chuyển đổi này được gọi là embedding ( $g_2$  và  $h_2$ ).

## 2 Math for PCA

Một gray-scale image được biểu diễn bởi ma trận  $X \in \mathbb{R}^{m \times n}$ .

Có thể chọn nhiều performance measure khác nhau để đánh giá độ tương quan giữa hai ảnh trong đó có thể kể tới hai measure performance sau:

- Inner product chỉ ra độ tương quan về hướng giữa hai ảnh. Measure này có nhược điểm là không xét được mối quan hệ giữa các điểm gần nhau (dù nhiều khả năng là các điểm này liên quan đến nhau). Tuy nhiên, đây vẫn là một measure hiệu quả trong nhiều trường hợp nên vẫn là một measure tốt.
- Norm chỉ ra khoảng cách giữa hai ảnh. Measure này khắc phục được nhược điểm nêu trên.

## 3 Principle Component Analysis (PCA)

Ta xấp xỉ  $X$  bởi các basis vectors  $(X_1, X_2, \dots, X_k)$ :

$$X \approx X_0 + a_1^X X_1 + a_2^X X_2 + \dots + a_k^X X_k = \tilde{X}_k$$

PCA tìm  $\tilde{X}_k$  xấp xỉ tốt nhất  $X$  hay  $\sum_{t=1}^D \|X_t^k - \tilde{X}_t^k\|$  đạt giá trị nhỏ nhất. Rõ ràng rằng  $k$  càng lớn thì sự xấp xỉ càng chính xác nhưng tốc độ chậm hơn.

Keyword để xem kĩ hơn về toán : SVD, covariance matrix