

Bài 4 :

Học có giám sát :

Là bài toán với bộ dữ liệu đầu vào có dạng (X,Y) trong đó X gồm các điểm dữ liệu mang đặc trưng và Y là giá trị đầu ra của từng điểm dữ liệu. Học có giám sát tìm ra các tham số mô hình sao cho sai số giữa đầu ra dự đoán và đầu ra thực tế là đủ nhỏ.

VD : Các bài toán phân loại có nhãn và hồi quy.

- Với bài toán dự đoán giá vàng trong tương lai.

X : các đặc trưng : Lãi suất ngân hàng, GDP của người dân, thời điểm,.....

Y : Giá vàng tại từng thời điểm.

Từ bộ dữ liệu trong quá khứ, xây dựng mô hình với tham số θ sao cho

$Y_{\text{predict}} = X * \theta$ và Y có sai số đủ nhỏ.

- Bài toán phân loại chữ số viết tay.

X : Bức ảnh kích thước 64×64 .

Y : Nhãn đối với từng bức ảnh.

Học không giám sát :

Là bài toán với bộ dữ liệu chỉ gồm X là các đặc trưng của dữ liệu. Thuật toán sẽ dựa vào cấu trúc của dữ liệu để thực hiện một công việc nào đó.

VD : Các bài toán liên quan đến clustering.

- X : Tập dữ liệu người dùng trên facebook. Mỗi một điểm X_i tương ứng với 1 người dùng. Phân cụm tập dữ liệu X thành k tập sao cho các người dùng trong mỗi tập có sự tương đồng với nhau và khác biệt so với các tập khác.

