

# Machine Learning

Lecturer: Doctor Bui Thanh Hung

Data Science Laboratory

Faculty of Information Technology

Industrial University of Ho Chi Minh city

Email: [hung.buithanhcs@gmail.com](mailto:hung.buithanhcs@gmail.com) ([buithanhhung@iuh.edu.vn](mailto:buithanhhung@iuh.edu.vn))

Website: <https://sites.google.com/site/hungthanhbui1980/>

## Learning là quá trình được trình bày trong hình sau:

- **Given:**

- Instances  $X$ : Possible days, each described by the attributes *Sky*, *AirTemp*, *Humidity*, *Wind*, *Water*, *Forecast*

- Target function  $c$ :  $EnjoySport : X \rightarrow \{0, 1\}$

- Hypotheses  $H$ : Conjunctions of literals. E.g.

$\langle ?, Cold, High, ?, ?, ? \rangle$ .

- Training examples  $D$ : Positive and negative examples of the target function

$\langle x_1, c(x_1) \rangle, \dots \langle x_m, c(x_m) \rangle$

- **Determine:** A hypothesis  $h$  in  $H$  such that  $h(x) = c(x)$  for all  $x$  in  $D$ .

Bạn hãy thực hiện các bài tập sau:

### Bài 1:

Mẫu dưới đây thu nhận ngẫu nhiên độ tuổi của người mắc bệnh tiểu đường ở một địa bàn trong thời gian khảo sát:

35 37 40 42 55 57 60 61 66 45

Có ý kiến cho rằng "Người mắc bệnh tiểu đường có độ tuổi trung bình trên 45 tuổi". Với mức ý nghĩa 5% bạn hãy kiểm định ý kiến này đúng hay sai? Hãy tính toán và đưa ra kết luận?

## Bài 2:

Điều tra chi tiêu hàng năm của 40 công nhân khu công nghiệp A thấy trung bình mẫu là 34 triệu đồng và độ lệch chuẩn mẫu là 5,8 triệu đồng. Điều tra công nhân tại khu công nghiệp B thì chi tiêu trung bình là 35 triệu đồng và độ lệch chuẩn là 5 triệu đồng.

- Với mức ý nghĩa 5%, có thể cho rằng chi tiêu trung bình của công nhân khu công nghiệp A và B là như nhau không?

- Tính p-value của kiểm định và cho biết kết luận thu được có thể mắc phải sai lầm loại mấy? Giả thiết chi tiêu công nhân là biến ngẫu nhiên phân phối chuẩn.

## Bài 3:

Quảng cáo của một công ty cho rằng tỷ lệ phế phẩm của công ty trên thị trường tối đa là 12%. Kiểm tra 400 sản phẩm của công ty trên thị trường thì có 60 phế phẩm. Với mức ý nghĩa 5% có thể cho rằng quảng cáo của công ty là đúng không?