Cuối kỳ Lập trình phân tích dl 1

# Tuần 6

***Học máy***

*Dựa trên Histonical data*

1. Dự báo (Prediction)

Ví dụ: Dự báo ngày mai nhiệt độ là bao nhiêu

1. Phân lớp (Classification)

Vd: ngày mai trời nắng hay mưa

Vd: tin nào tin giả tin nào tin thật

1. Gom cụm (Clustering)
2. Luật kết hợp (Association Rule)

Vvdaps dụng nhiueef cho thương mại điện tử (mua trứng mua rau và đề xuất ra mình cần mua gì tiếp theo)

* 3 phương pháp kỹ thuật chính
* Học có giám sát(Supervised)
* Học không giám sát (un-supervised)
* Học bán giám sát(Semi-Supervised)
* Phụ thuộc vào đầu ra và đầu vào (dựa vào tập dữ liệu)

giã sử có 3 cột điểm

|  |  |  |  |
| --- | --- | --- | --- |
| Histonical Data |  |  |  |
| T | V | Thi | Pvan |
| 8.5 | 7.0 | 6.5 | No |
| 7.5 | 6.0 | 8.0 | Yes |
| 8.0 | 8.0 | 4.0 | Yes |

VD1 dự báo điểm thi dựa vòa điểm T và Văn

Input Điểm T và V

Output điểm thi Là những con số thực nên bài toán**( dự báo**)(prediction)

* Học có giám sát input và ouput có trong tập dữ liệu

Vd 2 Dự báo đậu hay rớt PV dựa trên T và Văn

Input DIểm T V

Output phỏng vấn(Giá trị Yes/No ) là gtri định tính nên bài toán **Phân lớp**

* Học có giám sát là input và output có trong tập DL
* Học không giám sát khi có input trong Tập DL nhưng không có output trong tập Dl

VD dựa vào toán và văn và thi đưa chúng vào 2 nhóm khác nhau. Không có cái output

* Bán giám sát là input có nhưng output không đầy đủ hoặc không có luôn ( cách làm là điểu chế tập DL output)

**Cách giải quyết bài toán**

1. Xác định bài toán
2. Có dữ liệu
3. Xử lý dữ liệu
4. DL có gì Input và Output là cái gì 🡺 chọn mô hình học máy 🡺 lựa chọn mô hình để giải quyết bài toán

Dữ liệu đầu vào input 🡺 Model 🡺 giải quyết được Output

Đánh giá cái model tốt hay không tốt

* Đầu tiên có input rồi tách nó ra thành 2 phần(Crass Valiation)

+ Train (80% dữ liệu): dựa vào tk này để kiếm mô hình 🡺 kiếm được mô hình và đánh giá mô hình dựa vào 80%(học 80%)

+ Test(20%) đi kiểm tra 20% còn lại (Ktra 20%)

Nếu đánh giá tốt thì bước tiếp theo là Triển khai sản phẩm, và đưa ra thực tế. tiếp tục thu thập Feedback của kahchs hàng

# Tuần 7

Biến đổi dữ theo dạng minmax scale dể dữ thiệu theo khoảng [0.1] cho nó cùng một khoảng dữ liệu