

Project Giữa kỳ

Môn: Nhập môn Học Máy (Introduction to Machine Learning)

1. Yêu cầu

Mục tiêu: Xây dựng các mô hình học máy cho bài toán dự đoán:

- 1) (1 điểm) Mô tả dữ liệu, vẽ đồ thị về dữ liệu
- 2) (2 điểm) Chuẩn hoá dữ liệu trước khi đưa vào mô hình, xử lý missing data nếu có
- 3) (3 điểm) Xây dựng các mô hình học máy cho bài toán Classification hoặc Regression, thử nghiệm các mô hình sau:
kNN, NB, SVM
- 4) (1 điểm) Tự bổ sung thêm 1 mô hình nào tự tìm hiểu (Decision Tree, Neural Networks, ...)
- 5) (1 điểm) Đánh giá độ chính xác qua các độ đo precision, recall, F-score, accuracy
- 6) (2 điểm) Tìm hiểu về vấn đề Overfitting và trong chương trình trên đã áp dụng như thế nào? (phần này cần trình bày bằng slide)

2. Yêu cầu về Dữ liệu:

- 1) Lấy dữ liệu từ UCI, theo link sau:
<https://archive.ics.uci.edu/ml/datasets.php?format=&task=&att=mix&area=&numIns=&type=mvar&sort=attTypeDown&view=table>
- 2) Yêu cầu dữ liệu tại phải có attribute gồm nhiều kiểu dữ liệu khác nhau, giống như sau:

Attribute Types	
	Categorical, Real, Integer
	Categorical, Real
	Categorical, Real

- 3) Dữ liệu không được trùng với dữ liệu đã làm trong các bài tập của các buổi học. Dữ liệu không được trùng với các bạn khác trong cùng nhóm.

3. Nộp bài

Cho các kết quả vào thư mục, nén lại và nộp. Tên thư mục là Họ tên và MSSV

- Chương trình: file họ_tên_mssv.jpynb
 - o Nếu có nhiều chương trình thì bổ sung thông tin Model vào tên file
- Slide:

