

Bài tập 4

Bài 1: Tải về file winequality.csv về các số đo của rượu vang và chất lượng của rượu
Liên kết: <http://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-red.csv>

- Đây là bộ data của đại học California-Berkeley
- Bộ data gồm 1599 mẫu rượu vang, mỗi mẫu gồm 11 loại chỉ số và đánh giá của chuyên gia về chất lượng rượu (cột quality, điểm số từ 0 đến 10)

Chú ý:

- Dữ liệu sử dụng dấu chấm phẩy (;) để ngăn giữa các cột
- Tên các cột có chứa dấu cách

In ra dữ liệu vừa tải về, ý nghĩa các cột thuộc tính

- fixed acidity Nồng độ axit tartaric
- volatile acidity Tính axit
- citric acid Nồng độ axit Citric
- residual sugar Nồng độ đường dư
- chlorides Nồng độ clo
- free sulfur dioxide Nồng độ acid sulfuric tự do
- total sulfur dioxide Nồng độ acid sulfuric
- density Mật độ (khối lượng/đơn vị thể tích)
- pH Độ pH
- sulphates Nồng độ sunfat
- alcohol Nồng độ chất alcohol

a) Sử dụng các cột thuộc tính “alcohol” để tìm tương quan giữa thuộc tính này và điểm chất lượng rượu

b) Sử dụng tất cả 11 thuộc tính để tìm tương quan giữa các thuộc tính với điểm chất lượng rượu

Bài 2 Hãy thực hiện các công việc sau:

a) Hãy tạo một bộ dữ liệu có file “K59.csv” có các cột sau: soluongmau, KNN, FIS.
Có nội dung như sau:

soluongmau	KNN	FIS
300	0.904	0.912
500	0.893	0.901
700	0.863	0.863
1000	0.823	0.812
1200	0.802	0.794
1500	0.783	0.743
1800	0.765	0.732
2000	0.752	0.723

- b) Hãy đọc file “K58.csv” đã tạo ở trên, in nội dung ra màn hình
- c) Hãy vẽ biểu đồ đường so sánh sự tương quan giữa 2 phương pháp KNN, FIS.
- d) Xuất biểu đồ đã vẽ ở trên ra file “K58.png”
- e) Hãy tìm hệ số tương quan hồi quy và sai số giữa phương pháp KNN với phương pháp FIS.

Bài 3. Cho bộ dữ liệu về Tỷ giá ngoại tệ mua vào so với đồng VNĐ như sau:

Mã NT	AUD	CAD	CHF	EUR	GBP	HKD	JPY	SGD	THB	USD
Mua tiền mặt	16.815	17.138	22.568	26.172	29.876	2.864	201	16.734	692	22.775

- Lưu bộ dữ liệu trên ở file `tigia.csv`, có 3 cột: Mã NT, Mua tiền mặt
- Đọc file `tigia.csv` đã tạo ở câu 1, in nội dung dữ liệu đọc được ra màn hình
- Sắp xếp Mã NT theo thứ tự Alphabet và lưu vào file `tigia2.csv`
- Vẽ biểu đồ cột chiều đứng thể hiện tương quan về tỉ giá của các ngoại tệ với VNĐ, yêu cầu thêm chú thích phù hợp cho các dòng và cột trên biểu đồ để có thể truyền đạt nguyên vẹn các thông tin đến người sử dụng biểu đồ.
- Lưu biểu đồ đã vẽ ở Câu d ra file `tigia.png`

Bài 4. Anh chị lựa chọn cho mình ba trong số các mô hình sau:

- Hồi quy tuyến tính
- Phân lớp dựa trên Naïve bayes
- Phân lớp dựa trên SVM
- Phân cụm dữ liệu cứng K-mean
- Phân cụm dữ liệu mờ FCM

Cài đặt thử nghiệm trên 2 bộ dữ liệu lấy từ bộ dữ liệu chuẩn của học máy:

<https://archive.ics.uci.edu/ml/datasets.html>

Bài 5. Anh chị lựa chọn cho mình 1 bài toán thực tế.

- Thu thập dữ liệu về bài toán đó.
- Thực hiện cài đặt một mô hình học máy dữ lý dữ liệu trên.