

TOÁN ỨNG DỤNG VÀ THỐNG KÊ

Đồ án Linear Regression

Nội dung đồ án

File "**wine.csv**" là cơ sở dữ liệu đánh giá chất lượng của 1200 chai rượu vang theo thang điểm 1 - 10 dựa trên 11 tính chất khác nhau.

```
In [1]: import pandas as pd
df = pd.read_csv('wine.csv')
```

```
In [2]: df
```

```
Out[2]:
```

	"quality"	fixed acidity	"volatile acidity"	"citric acid"	"residual sugar"	"chlorides"	"free sulfur dioxide"	"total sulfur dioxide"	"density"	"pH"	"sulfur dioxide"
0	5	7.4	0.700	0.00	1.9	0.076	11.0	34	0.99780	3.51	
1	5	7.8	0.880	0.00	2.6	0.098	25.0	67	0.99680	3.20	
2	5	7.8	0.760	0.04	2.3	0.092	15.0	54	0.99700	3.26	
3	6	11.2	0.280	0.56	1.9	0.075	17.0	60	0.99800	3.16	
4	5	7.4	0.700	0.00	1.9	0.076	11.0	34	0.99780	3.51	
...
1194	6	7.0	0.745	0.12	1.8	0.114	15.0	64	0.99588	3.22	
1195	6	6.2	0.430	0.22	1.8	0.078	21.0	56	0.99633	3.52	
1196	6	7.9	0.580	0.23	2.3	0.076	23.0	94	0.99686	3.21	
1197	6	7.7	0.570	0.21	1.5	0.069	4.0	9	0.99458	3.16	
1198	6	7.7	0.260	0.26	2.0	0.052	19.0	77	0.99510	3.15	

1199 rows × 12 columns



Xây dựng mô hình đánh giá chất lượng rượu sử dụng phương pháp hồi quy tuyến tính.

- Sử dụng toàn bộ 11 đặc trưng đề bài cung cấp.
- Sử dụng duy nhất 1 đặc trưng cho kết quả tốt nhất. (Gợi ý: Phương pháp Cross Validation)
- Xây dựng một mô hình của riêng bạn cho kết quả tốt nhất.

Lưu ý: sinh viên phải sử dụng thuật toán OLS đã được hướng dẫn trong phần bài tập. Sinh viên không được dùng các hàm có sẵn của các thư viện để thực hiện các tính toán.

Quy định bài nộp

- Thực hiện toàn bộ bài làm trên 1 tập tin Jupyter Notebook (.ipynb)
- Đầu bài phải có phần trình bày (bằng Markdown) thông tin sinh viên và giải thuật
- Cuối bài phải có phần mô tả ý tưởng thực hiện và mô tả các hàm

Quy định chấm bài

Những trường hợp sau đây sẽ bị 0 điểm toàn bộ đồ án:

- Nộp sai quy định bài nộp
- Thực thi mã nguồn báo lỗi

****LƯU Ý: SAO CHÉP BÀI LÀM CỦA NHAU SẼ BỊ 0 ĐIỂM TOÀN BỘ PHẦN THỰC HÀNH****