

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**ĐHQG TPHCM**

**Báo Cáo Đồ Án 3**

**Instructors: Nguyễn Đình Thúc**

**Nguyễn Văn Quang Huy**

**Nguyễn Trọng Hiến**

**Võ Nam Thục Đoan**

**Thông tin Cá Nhân:**

1. Tôn Thất Bách MSSV:20127442

Lưu ý: các đoạn chương trình cung cấp trong báo cáo có thể cần có hàm phụ để chạy được, xem chi tiết hơn ở mã nguồn gửi kèm báo cáo.

Mục lục

[I. Nội dung thực hiện 3](#_Toc109149971)

[*1. Sử dụng phương pháp hồi quy tuyến tính với 11 đặc trưng đề bài cung cấp để xây dựng mô hình đánh giá chất lượng rượu 3*](#_Toc109149972)

[*2. Sử dụng phương pháp Cross Validation để tìm ra đặc trưng tốt nhất từ đó xây dựng mô hình tương ứng 4*](#_Toc109149973)

[*3. Xây dựng mô hình riêng của bản thân cho ra kết quả tốt nhất 6*](#_Toc109149974)

[II. Tham khảo 9](#_Toc109149975)

# I. Nội dung thực hiện

1. **Sử dụng phương pháp hồi quy tuyến tính với 11 đặc trưng đề bài cung cấp để xây dựng mô hình đánh giá chất lượng rượu** 
   * Input: Mảng 2 chiều các số liệu từ file đề bài
   * Output: Mô hình theo yêu cầu đề bài
   * Ý tưởng thực hiện: Từ file đề bài **“wine.csv”** ta có:

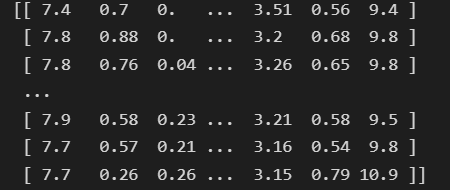
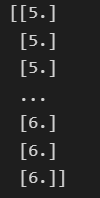
**Các biến độc lập:**

X={ "fixed acidity";"volatile acidity";"citric acid";"residual sugar";"chlorides";"free sulfur dioxide";"total sulfur dioxide";"density";"pH";"sulphates";"alcohol"}

**Biến Phụ Thuộc:**

Y={“quality”}

F1(x) = [“fixed acidity”] …, F11(x)=[“alcohol”]

Đặt A= [ F1(x), F2(x)… F11(x)] B=Y=[“quality”]

Theta=[Theta\_1] = (AT\*A)-1 \* AT \* B

[Theta\_2]

[….]

[Theta\_11]

Ảnh có chứa văn bản

Mô tả được tạo tự độngTa có mô hình theo yêu cầu:

* Cài đặt với Python  
  Ảnh có chứa văn bản

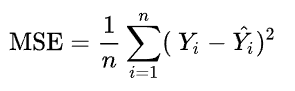
  Mô tả được tạo tự động

Ảnh có chứa văn bản

Mô tả được tạo tự độngKết quả mảng Theta với 11 đặc trưng và mô hình tương ứng:

Y=0.00593x1-1.10804x2-0.26305x3+0.01532x4-1.73050x5+0.00380x6-0.00390x7+4.33859x8-0.45854x9+0.72972x10+0.30886\*x11

1. Sử dụng phương pháp **Cross Validation** để tìm ra đặc trưng tốt nhất từ đó xây dựng mô hình tương ứng

* Ý tưởng thực hiện: **Cross Validation** có kết quả trả về là **độ sai số** của một đặc trưng trong việc xác định chất lượng rượu
* Độ sai số (MSE): là hàm dùng đánh giá chất lượng của một ước lượng hay dự báo cụ thể là đánh giá một đặc trưng trong việc đánh giá chất lượng rượu

Bước 1:Ta xáo trộn dữ liệu hiện tại bằng cách hoán đổi các dòng với nhau

Bước 2:Chia dữ liệu đã xáo trộn thành K tập hợp tùy ý (Mã nguồn đi kèm với báo cáo sử dụng K = 5) => Bộ dữ liệu có 5 tập hợp D1, D2, D3, D4, D5

Bước 3: Chọn đặc trưng cần kiểm tra độ sai số

Bước 3:Chọn từng tập hợp để làm tập hợp Test, các tập hợp còn lại sẽ là tập hợp Train VD: D1 là Test thì tập G={D2,D3,D4} là Train

Bước 4: Tìm mô hình tương ứng cho tập train với X={“Đặc trưng cần kiểm tra”}

Bước 5: Với đặc trưng vừa tìm được ta tìm Y\_Hat = Y(X)

Bước 6: Tìm MSE của Y và Y\_Hat

Sau đó ta đổi tập Test khác cho đến khi các tập hợp con đã được Test hết và trả về (Tổng các giá trị MSE)/(K tập hợp) ta sẽ có mức độ sai số của một đặc trưng

Lần lượt thử mức độ sai số của các đặc trưng và chọn đặc trưng có sai số ít nhất để tìm mô hình tối ưu

* Ảnh có chứa văn bản

  Mô tả được tạo tự độngCài đặt với Python

Mức sai số tương ứng với từng đặc trưng từ đó chọn ra đặc trưng tốt nhất và tìm mô hình

Ảnh có chứa văn bản

Mô tả được tạo tự động

1. Xây dựng mô hình riêng của bản thân cho ra kết quả tốt nhất
   * Ý tưởng thực hiện:

Sử dụng độ sai số từ phương pháp **Cross Validation** ở câu trước để đánh giá chất lượng của mô hình bản thân đưa ra

Mô hình 1:

Mã nguồn:

Ảnh có chứa văn bản

Mô tả được tạo tự động

Kết quả thực thi mã nguồn:

Ảnh có chứa văn bản

Mô tả được tạo tự động**Sai số của Mô hình 1 là : 0.49365**

Mô hình 2:

Ảnh có chứa văn bản, thiết bị đo, thiết bị

Mô tả được tạo tự động

Mã nguồn:

Ảnh có chứa văn bản

Mô tả được tạo tự động

Ảnh có chứa văn bản

Mô tả được tạo tự độngKết quả thực thi mã nguồn:

**Mô hình 2 có độ sai số là 0.4884**

***Kết luận:***

# Ảnh có chứa văn bản Mô tả được tạo tự độngII. Tham khảo

<https://vi.wikipedia.org/wiki/Ki%E1%BB%83m_ch%E1%BB%A9ng_ch%C3%A9o>

Các tài liệu được cung cấp trên Moodle

<https://vi.wikipedia.org/wiki/Sai_s%E1%BB%91_to%C3%A0n_ph%C6%B0%C6%A1ng_trung_b%C3%ACnh>