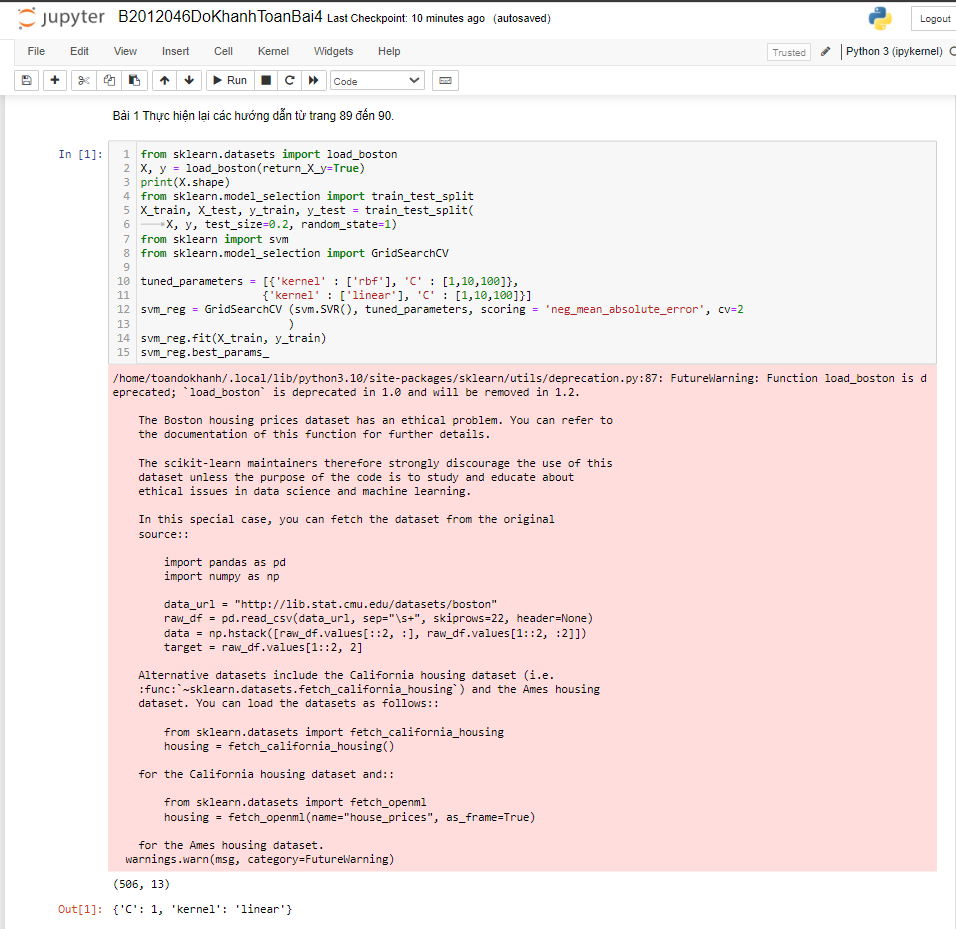
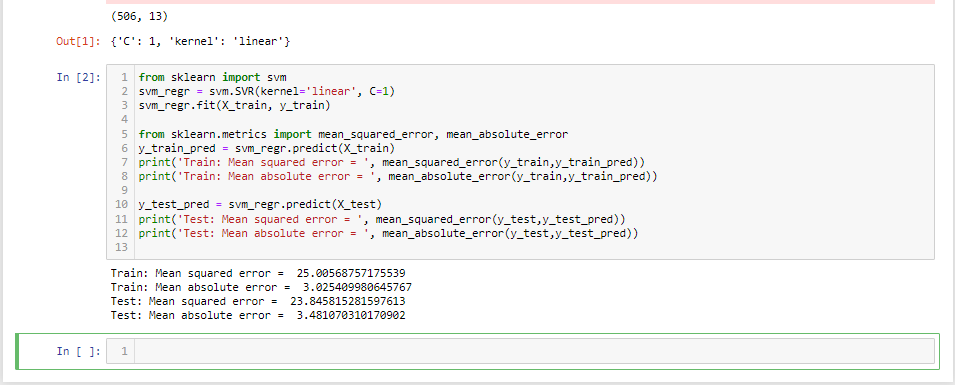
**Bài tập thực hành 05:**

Lựa chọn siêu tham số

1. Thực hiện lại các hướng dẫn từ trang 89 đến 90.

**Bài làm**

****

****

1. Lựa chọn siêu tham số để làm gì? Nêu các phương pháp để lựa chọn siêu tham số[[1]](#footnote-0).

**Bài làm**

Lựa chọn siêu tham số là quá trình tìm ra các giá trị tối ưu cho các biến cấu hình bên ngoài của một mô hình máy học. Các siêu tham số ảnh hưởng đến kiến trúc, chức năng và hiệu năng của mô hình. Một số phương pháp để lựa chọn siêu tham số là:

Tìm kiếm thủ công: Thử nghiệm thủ công với các bộ siêu tham số ngẫu nhiên và chọn bộ tốt nhất dựa trên một tiêu chí đánh giá, chẳng hạn như hàm mất mát hoặc độ chính xác.

Tìm kiếm theo lưới: Xây dựng một mô hình cho mỗi hoán vị của tất cả các giá trị siêu tham số được chỉ định trong một lưới và chọn mô hình tốt nhất.

Tìm kiếm ngẫu nhiên: Chọn một tổ hợp ngẫu nhiên của các siêu tham số từ không gian tìm kiếm và lặp lại quá trình này với số lần cố định. Tìm kiếm ngẫu nhiên có thể hiệu quả hơn tìm kiếm theo lưới khi không gian siêu tham số lớn và không đồng đều.

Tìm kiếm Bayes: Sử dụng các quan sát trước đó về hiệu năng của các bộ siêu tham số để xây dựng một mô hình xác suất và dự đoán bộ siêu tham số tiếp theo có khả năng cao nhất cải thiện hiệu năng. Tìm kiếm Bayes có thể khai thác được các khu vực có triển vọng trong không gian tìm kiếm và tiết kiệm thời gian đào tạo.

Link tham khảo:

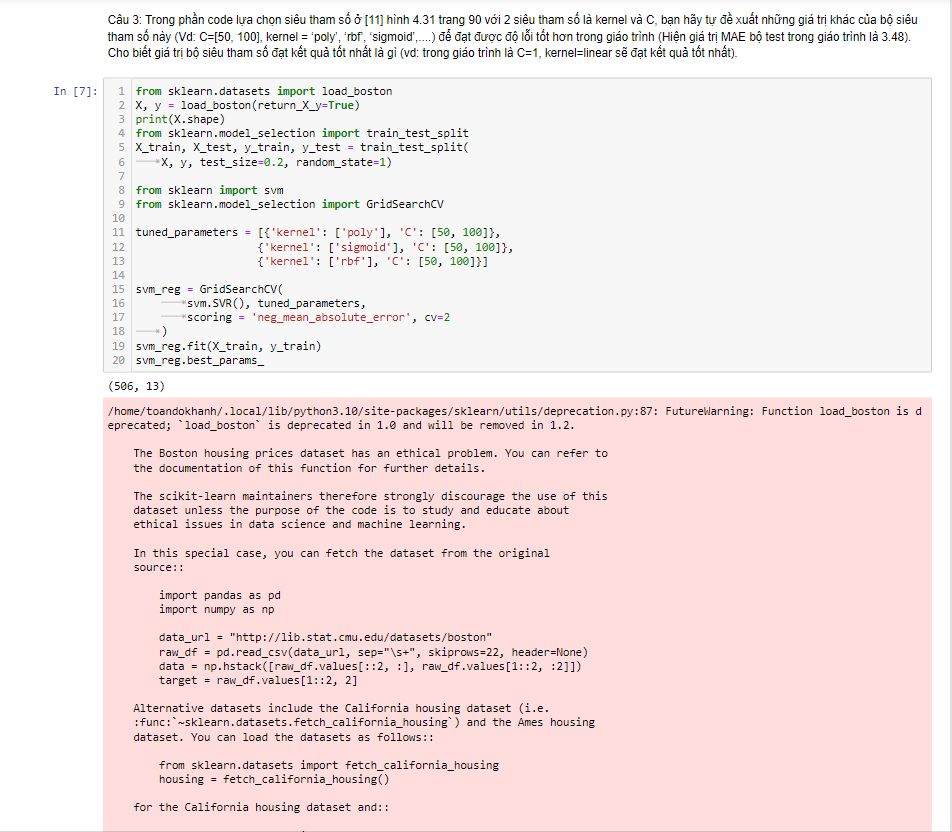
<https://ichi.pro/vi/4-ky-thuat-dieu-chinh-sieu-tham-so-160858093703577>

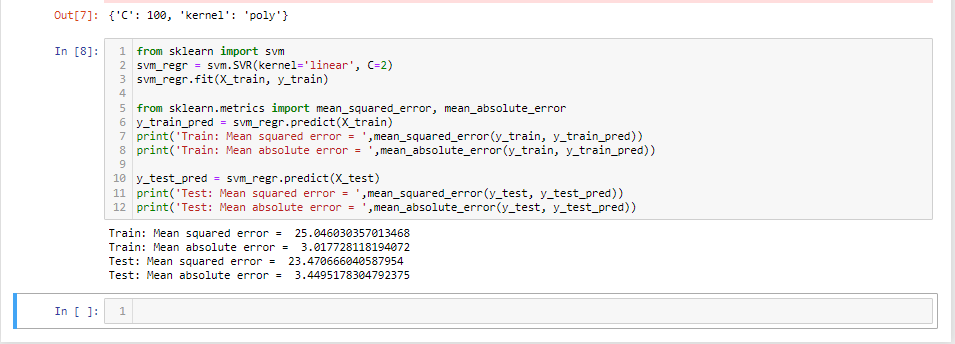
<https://aws.amazon.com/vi/what-is/hyperparameter-tuning/>

<https://www.noron.vn/post/cach-chon-sieu-tham-so-cho-mang-neuron-qlk4scudu65>

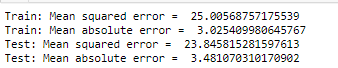
1. Trong phần code lựa chọn siêu tham số ở [11] hình 4.31 trang 90 với 2 siêu tham số là kernel và C, bạn hãy tự đề xuất những giá trị khác của bộ siêu tham số này (Vd: C=[50, 100], kernel = ‘***poly’, ‘rbf’, ‘sigmoid’,....)*** để đạt được độ lỗi tốt hơn trong giáo trình (Hiện giá trị MAE bộ test trong giáo trình là 3.48). Cho biết giá trị bộ siêu tham số đạt kết quả tốt nhất là gì (vd: trong giáo trình là C=1, kernel=linear sẽ đạt kết quả tốt nhất).

**Bài làm**

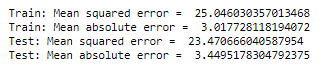
****

****

Kết quả trong giáo trình



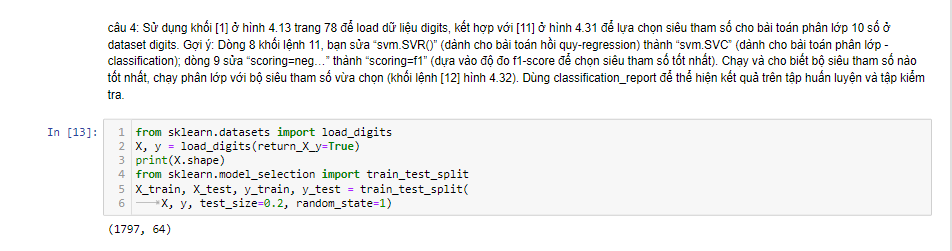
Kết quả khi đổi tham số C=[50, 100], kernel = ‘***poly’, ‘rbf’, ‘sigmoid’***

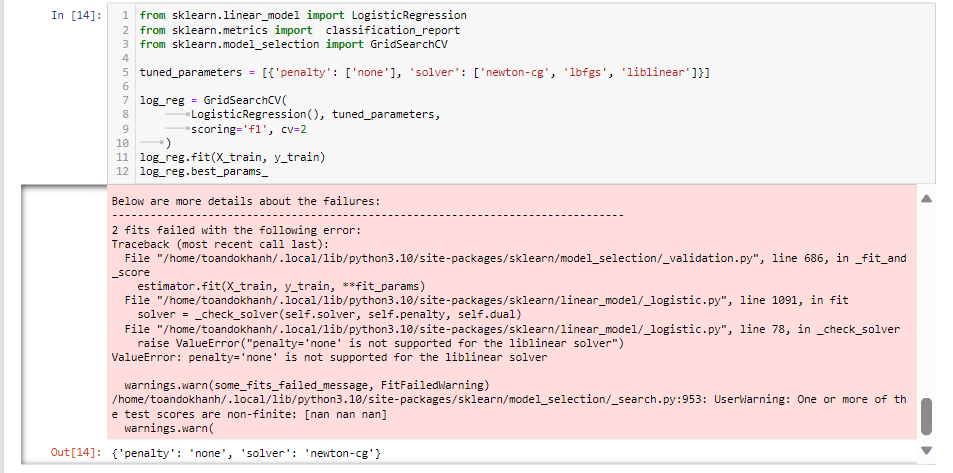


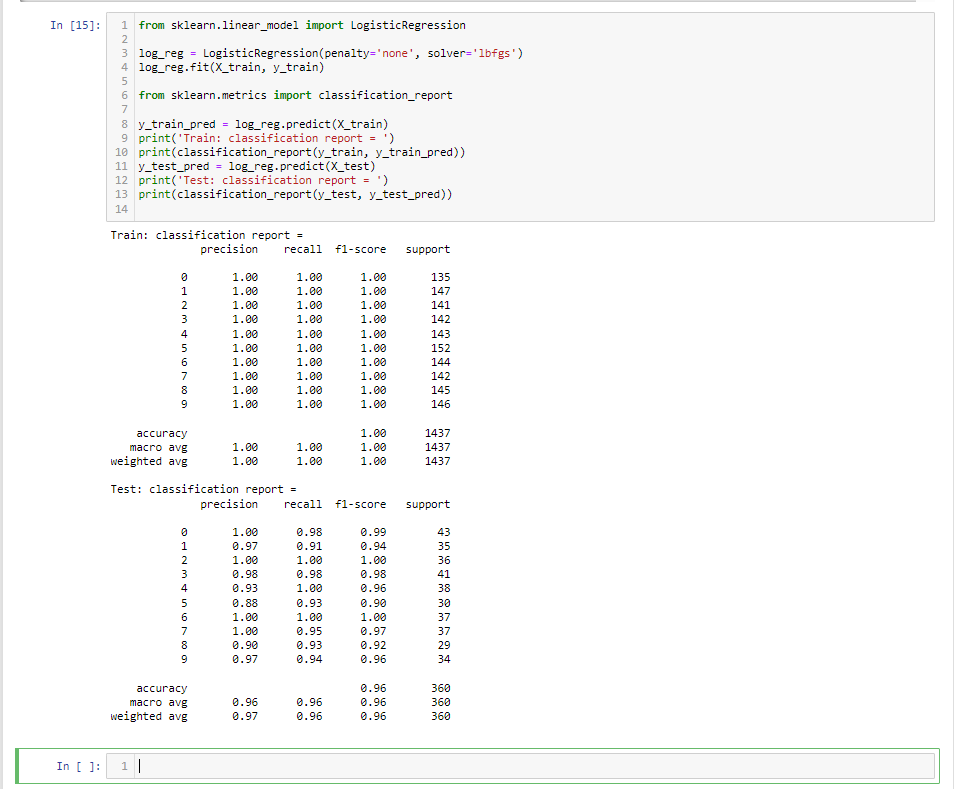
=> Dựa trên kết quả của hai lần chạy SVM với các tham số tinh chỉnh khác nhau, có thể nhận thấy rằng chạy thứ hai (Tuned Parameters 2) với kernel 'rbf' và C = 100 cho thấy hiệu suất tốt hơn trên tập kiểm tra, có MSE và MAE thấp hơn so với chạy đầu tiên (Tuned Parameters 1).

1. Sử dụng khối [1] ở hình 4.13 trang 78 để load dữ liệu digits, kết hợp với [11] ở hình 4.31 để lựa chọn siêu tham số cho bài toán phân lớp 10 số ở dataset digits. Gợi ý: Dòng 8 khối lệnh 11, bạn sửa “svm.SVR()” (dành cho bài toán hồi quy-regression) thành “svm.SVC” (dành cho bài toán phân lớp - classification); dòng 9 sửa “scoring=neg…” thành “scoring=f1” (dựa vào độ đo f1-score để chọn siêu tham số tốt nhất). Chạy và cho biết bộ siêu tham số nào tốt nhất, chạy phân lớp với bộ siêu tham số vừa chọn (khối lệnh [12] hình 4.32). Dùng classification\_report để thể hiện kết quả trên tập huấn luyện và tập kiểm tra.

**Bài làm**

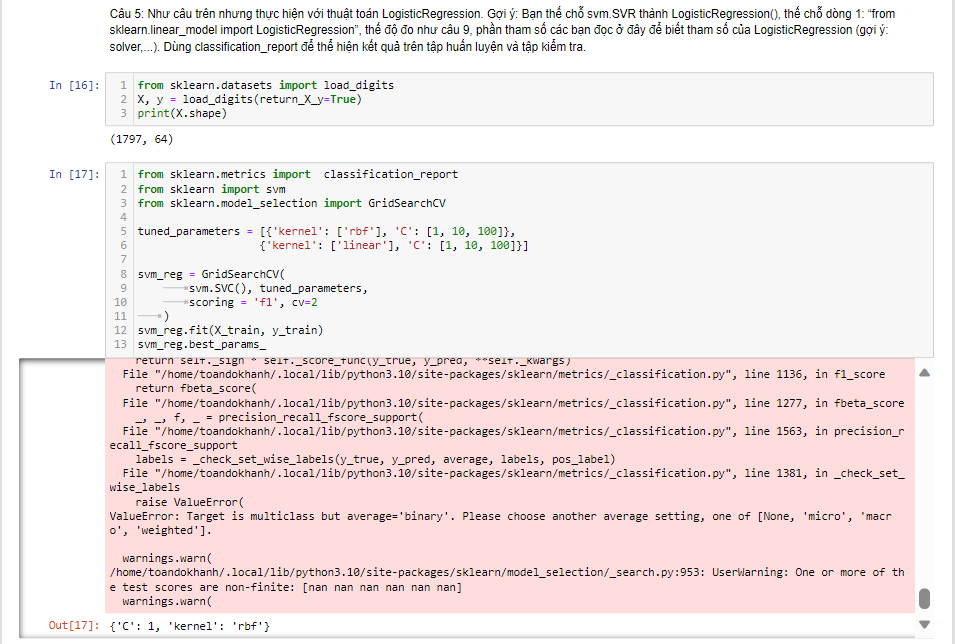
****

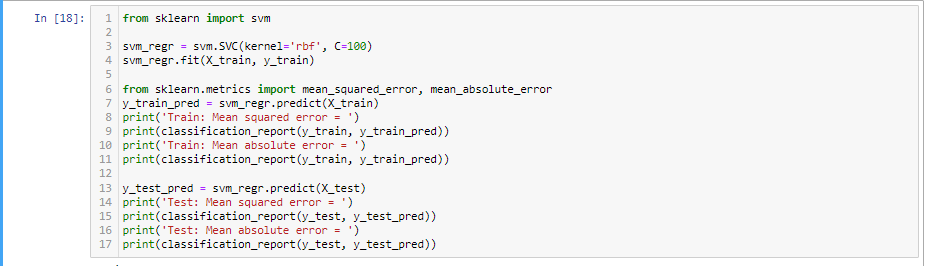
****

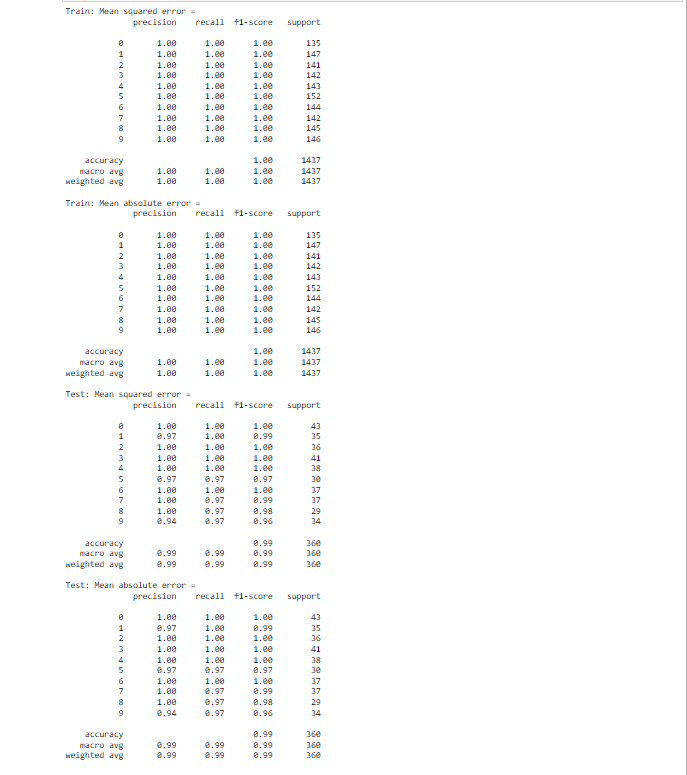
****

1. Như câu trên nhưng thực hiện với thuật toán LogisticRegression. Gợi ý: Bạn thế chỗ svm.SVR thành LogisticRegression(), thế chỗ dòng 1: “from sklearn.linear\_model import LogisticRegression”, thế độ đo như câu 9, phần tham số các bạn đọc ở đây[[2]](#footnote-1) để biết tham số của LogisticRegression (gợi ý: *solver,...)*. Dùng classification\_report để thể hiện kết quả trên tập huấn luyện và tập kiểm tra.

**Bài làm**







1. https://scikit-learn.org/stable/modules/grid\_search.html [↑](#footnote-ref-0)
2. https://scikit-learn.org/stable/modules/generated/sklearn.linear\_model.LogisticRegression.html [↑](#footnote-ref-1)