

Bài tập về nhà

Buổi số: 04

Bài 1: Support Vector Machine (SVM) là một trong những thuật toán phân lớp hiệu quả và linh động. Dựa trên những kiến thức đã học trên lớp và đọc thêm tài liệu ở nhà, em hãy trình bày ngắn gọn những hiểu biết của mình về SVM.

Bài 2¹: *(Thực hành với thư viện scikit-learn)*

Thư viện scikit-learn cung cấp các lớp như SVC, NuSVC và LinearSVC có khả năng thực hiện phân loại nhị phân và đa lớp trên một tập dữ liệu. Moons dataset² là một tập dữ liệu (toy dataset) thường được dùng để giúp trực quan hóa các thuật toán phân lớp. Em hãy lựa chọn một trong các lớp của thư viện scikit-learn để lập trình giải bài toán phân lớp sử dụng moon dataset. Đầu ra có thể vẽ đường bao phân lớp để giúp trực quan hóa việc phân lớp cũng như ưu điểm của thuật toán mà em đã thực hiện.

Bài 3: *(Giải bài toán bằng cách lập trình)*

Sử dụng Python, lập một chương trình để giải quyết bài toán phân lớp trong ngôi nhà thông minh đã làm trên lớp với tập dữ liệu huấn luyện 14 mẫu. Người dùng nhập vào giá trị một mẫu mới: (21°C, 52%). Chương trình in ra kết quả phân lớp và vẽ hình minh họa kết quả. Trên hình minh họa kết quả, các em vẽ các điểm trong tập dữ liệu huấn luyện với màu xanh nước biển+hình tròn cho Bedroom và màu đỏ+hình vuông cho Living room; mẫu mới dùng màu xanh lá cây(+hình vuông hoặc hình tròn tùy thuộc vào kết quả phân lớp).

¹ Các em có thể tham khảo thêm về SVM trong scikit-learn <https://scikit-learn.org/stable/modules/svm.html>

² Tham khảo về việc tạo tập dữ liệu moon dataset tại https://scikit-learn.org/stable/modules/generated/sklearn.datasets.make_moons.html