Tìm hai tập dữ liệu có nhãn. Nhớ chia tập dữ liệu thành các tập huấn luyện/kiểm tra.

1. Chạy k-nn trên hai tập dữ liệu này.

2. Tính lỗi phân loại, độ chính xác, độ thu hồi và điểm f1 (bằng cách so sánh các nhãn lớp thu được với nhãn dự đoán và nhãn gốc của dữ liệu kiểm tra).

3. Thay đổi giá trị k và nhận xét về kết quả.

4. Thử chuẩn hóa tập dữ liệu đầu vào. Hiệu suất có tốt hơn không?

5. Áp dụng PCA và SVD trên tập dữ liệu, vậy hiệu suất của k-nn trên dữ liệu dự đoán mới là bao nhiêu? Giải thích câu trả lời.

1. Lặp lại các câu hỏi này cho SVM.



