

Khi tạo bộ phân lớp thứ tư, bằng cách lấy mẫu, ta tạo ra tập huấn luyện cho bộ phân lớp này : $D4 = \{ X_1, X_3, X_8, X_9, X_{10}, X_{11} \}$

Khi tạo bộ phân lớp thứ năm, bằng cách lấy mẫu, ta tạo ra tập huấn luyện cho bộ phân lớp này : $D5 = \{ X_1, X_3, X_6, X_7, X_{12}, X_{13}, X_{14} \}$

Khi tạo bộ phân lớp thứ sáu, bằng cách lấy mẫu, ta tạo ra tập huấn luyện cho bộ phân lớp này : $D6 = \{ X_1, X_3, X_6, X_7, X_{10}, X_{11} \}$

Các bộ phân lớp thành phần đều là bộ phân lớp 1-lần cận gần nhất.

Xét mẫu thử (3.5, 2.8). Bộ phân lớp tổ hợp sẽ quyết định mẫu thử thuộc lớp nào ?

7. (1.5 điểm)

- a. Trong bài toán phát hiện xâm nhập mạng (intrusion detection), nếu dùng phương pháp dựa vào gom cụm (clustering-based) thay vì phương pháp dựa vào phân lớp (classification-based), thì hệ thống sẽ có được những lợi điểm gì? (0.5 điểm)
- b. Hãy giải thích hai thuật ngữ trong giải thuật gom cụm BIRCH: đặc trưng cụm (cluster feature) và cây đặc trưng cụm (CF tree). (0.75 điểm)
- c) (Đúng/Sai) BIRCH là giải thuật gom cụm gia tăng. (0.25 điểm)

8. (2 điểm)

- a. Nêu công dụng chính của kiến trúc nhiều tầng trong mạng nơ ron học sâu. (0.25 điểm)
- b. Trong lãnh vực mạng nơ ron học sâu, *mô hình sinh* (generative model) và *mô hình phân biệt* (discriminative model) khác biệt nhau như thế nào? (0.5 điểm)
- c. Nêu tên hai giải thuật dùng trong hai giai đoạn của quá trình huấn luyện mạng Deep Belief Network; một giai đoạn là học không giám sát và một giai đoạn là học có giám sát. (0.5 điểm)
- d. Giải thích công dụng của tầng tích chập (convolution) và tầng gộp (pooling) trong mạng nơ ron tích chập. (0.5 điểm)
- e. Hãy nêu tên giải thuật dùng để huấn luyện mạng nơ ron tích chập (CNN) (0.25 điểm)

9. (0.75 điểm) Ứng với các phần mềm mã nguồn mở sau đây để hỗ trợ cho lãnh vực học sâu, hãy nêu tên ngôn ngữ lập trình được dùng để hiện thực mỗi phần mềm đó.

TensorFlow

Keras

PyTorch