Đề thi cuối kỳ học kỳ I (2020-2021) Moân : Học Máy và Ứng dụng

Lớp Cao học Thôợi gian: 80 phuùt

(Sinh viên được phép tham khảo slide moân học)

Đề gồm 2 trang

1. (1 điểm) Trả lời các câu hỏi sau đây, mỗi câu 0.25 điểm

1.1 Cho một tập dữ liệu không khả tách một cách tuyến tính, bộ phân lớp nhị phân nào sau đây có thể tách tập dữ liêu ra làm hai lớp một cách đúng đắn:

A. SVM tuyến tính

B. Perceptron sử dụng perceptron training rule

C. SVM với RBF kernel

D. Các ba bộ phân lớp trên đều không thể.

1.2 Thư viện phần mềm học sâu nào sau đây có thể làm việc với môi trường mobile:

A. Keras

B. TensorFlow

C. Torch

D. Pytorch

- 1.3 (Đúng/Sai) SVM và ANN có thể được dùng cho cả hai bài toán phân lớp và hồi quy.
- **1.4** (Đúng/Sai) Trong mạng nơ ron tích chập (CNN) các tầng gộp (pooling layer) nhằm thu giảm độ phân giải không gian của hình ảnh khi phân lớp hình ảnh.
- **2.** (**0.5 điểm**) Các tiêu chí đánh giá ngoại như Rand, Jackard, v.v. để đánh giá chất lượng gom cụm được áp dụng trong trường hợp nào của gom cụm?
- **3.** (**0.5 điểm**) Nêu điểm khác biệt giữa giải thuật *suy giảm độ dốc* (gradient descent) và giải thuật *suy giảm độ dốc tăng dần* (incremental gradient descent) khi huấn luyện một đơn vị perceptron tuyến tính.
- **4. (0.75 điểm)** Khi xác định cấu hình của mạng nơ ron cho một bài toán phân lớp, căn cứ vào đâu để xác định số nút của tầng nhập và số nút của tầng xuất? Và bằng cách nào xác định số nút thích hợp cho tầng ẩn?

5. (1.5 điểm)

- a) Giải thích các thuật ngữ khoảng biên (margin), siêu phẳng tách với khoảng biên tối đa (maximum margin hyperplane), siêu phẳng hỗ trợ (support hyperplane), véc tơ hỗ trợ (support vector). (1 điểm)
- b. Giải thích ý nghĩa của các biến bù ξ_i và thông số C trong máy vec tơ hỗ trợ với khoảng biên mềm (soft margin) (0.5 diễm)
- **6.** (1.5 điểm) Cho tập mẫu gồm 14 mẫu hai chiều như sau:

 $X_1 = (1, 1, 1), X_2 = (1, 2, 1), X_3 = (2, 1, 1), X_4 = (2, 1.5, 1), X_5 = (3, 2, 1),$

 $X_6 = (4, 1.5, 2), X_7 = (4, 2, 2), X_8 = (5, 1.5, 2), X_9 = (4.5, 2, 2), X_{10} = (4, 4, 3),$

 $X_{11} = (4.5, 4, 3), X_{12} = (4.5, 5, 3), X_{13} = (4, 5, 3), X_{14} = (5, 5, 3),$

Mỗi mẫu được biểu diễn bởi đặc trưng 1, đặc trưng 2 và nhãn lớp. Ta xây dựng một tổ hợp bộ phân lớp bằng kỹ thuật bagging như sau:

Khi tạo bộ phân lớp thứ nhất, bằng cách lấy mẫu, ta tạo ra tập huấn luyện cho bộ phân lớp này : $D1 = \{X_2, X_4, X_5, X_8, X_9, X_{12}, X_{13}, X_{14}\}$

Khi tạo bộ phân lớp thứ hai, bằng cách lấy mẫu, ta tạo ra tập huấn luyện cho bộ phân lớp này : $D2 = \{ X_2, X_4, X_5, X_8, X_9, X_{10}, X_{11} \}$

Khi tạo bộ phân lớp thứ ba, bằng cách lấy mẫu, ta tạo ra tập huấn luyện cho bộ phân lớp này : $D3 = \{X_1, X_3, X_8, X_9, X_{12}, X_{13}, X_{14}\}$

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-lân 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 diễm)

- 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