

12/10/19

Đề thi cuối kỳ môn: Học máy

(Thời gian: 90 phút, mỗi câu nhỏ 1 điểm)

Câu I. 1) So sánh thuật toán xây dựng cây quyết định ID3 và C4.5

- 2) Dùng thuật toán ID3, xây dựng cây quyết định cho các đối tượng có hai đặc trưng, đặc trưng thứ nhất nhận hai giá trị 0 hoặc 1, đặc trưng thứ hai nhận giá trị trong 5 chữ cái. Dữ liệu quan sát được gồm hai lớp cho bởi bảng sau:

ω_1	(0,A)	(0,B)	(1,B)	(0,E)	(0,D)	(1,D)
ω_2	(1,A)	(0,C)	(1,C)	(1,E)		

Câu II. 1) Cho tập dữ liệu quan sát trong \mathbb{R}^2 gồm 3 lớp.

$$S_1 = \{(0,0), (1,0), (0,1)\}; S_2 = \{(1,1), (2,1)\}; S_3 = \{(-1,0), (-1,-1), (1,-2), (1,-1)\}$$

Anh/chị hãy chỉ ra các hàm phân biệt cho mỗi lớp khi dùng phương pháp khoảng cách cực tiểu theo khoảng cách Mahalanobis

- 2) Thiết kế mạng nơron MLP cho hàm $f: \mathbb{R}^7 \rightarrow \mathbb{R}^5$ nếu tập mẫu quan sát được gồm 100 mẫu: $D = \{(x^k, y^k)\}_{k=1}^{100}$

Câu III. Một địa phương có 420000 dân, phát hiện 60 người bị bệnh lạ. Trong số những người nhiễm bệnh, có 58 người bị sốt còn những người không nhiễm bệnh có 1% bị sốt. Có một người mới bị sốt, Anh /Chị hãy:

- Dùng tiêu chuẩn MAP để đoán xem người này có nhiễm bệnh không?
- Một liều thuốc điều trị sớm là 2 triệu đồng còn điều trị muộn là 200 triệu đồng, anh (chị) hãy quyết định xem có nên cho người này dùng thuốc không?

Câu IV.

- Trình bày phương pháp bagging xây dựng và kết hợp các bộ học.
- Một bộ phân lớp được kết hợp theo hình thức bỏ phiếu trọng số đều từ 5 bộ thành phần có kết quả đoán nhận tương đối độc lập với xác suất đúng của mỗi bộ là 0,7. Hãy ước lượng xác suất đúng của bộ kết hợp.

Câu V.

- Anh/chị hãy trình bày phương pháp học Q (Q-learning) trong học tăng cường cho bài toán đơn định
- So sánh hai thuật toán phân cụm k-centroid và k-mean.