

Question 2

Not yet answered

Marked out of 1.00

Cho tập dữ liệu gồm các điểm dữ liệu sau:  $A = (0; 2)$ ,  $B = (0; 1)$  và  $C = (1; 0)$ . Thuật toán K-means (với khoảng cách Euclid) với 2 cụm và tâm cụm được khởi tạo tại A và B. Khi hội tụ, tâm cụm sẽ là

Select one:

- ☐ a. A và C
- ☒ b. A và trung điểm BC
- ☐ c. C và trung điểm AB
- ☐ d. A và B

Time left 0:58:19

Question 3

Not yet answered

Marked out of 1.00

Trong mạng nơ-ron, đầu là lợi ích của hàm kích hoạt ReLU so với hàm kích hoạt Sigmoid?

Select one:

- ☒ a. ReLUs chấp nhận tính toán lan truyền ngược gradient nhanh hơn
- ☐ b. ReLUs có thể sử dụng cho tầng đầu ra, trong khi hàm sigmoid thì không thể
- ☐ c. ReLUs chấp nhận mô hình học một đường quyết định phi tuyến

Time left 0:58:09

Question 4

Not yet answered

Marked out of 1.00

Cây quyết định ...

Select one:

- ☐ a. không phải là phương pháp học có giám sát và cũng không phải là phương pháp học không giám sát
- ☐ b. là một phương pháp học có giám sát và cũng là một phương pháp học không giám sát
- ☒ c. là một phương pháp học có giám sát
- ☐ d. là một phương pháp học không giám sát

Time left 0:58:05

Question 5

Not yet answered

Marked out of 1.00

Một biến quan sát được có thể biểu diễn

Select one:

- ☐ a. bất kể thứ gì
- ☐ b. một sự kiện mà không thể quan sát được
- ☐ c. một sự kiện mà có thể quan sát được hoặc không
- ☒ d. một sự kiện mà có thể quan sát được trong thực tế

Time left 0:58:02

Question 6

Not yet answered

Marked out of 1.00

Hàm đánh giá nào để chọn thuộc tính kiểm tra là phù hợp nhất đối với các biến (thuộc tính) kiểu định danh có nhiều giá trị?

Select one:

- ☐ a. Entropy
- ☒ b. Gain Ratio
- ☐ c. Information Gain

Time left 0:57:59

Question 7

Not yet answered

Marked out of 1.00

Cho trước tập dữ liệu không có nhãn gồm N điểm dữ liệu:  $\{x_1, x_2, \dots, x_N\}$ . Chúng ta chạy K-means với 50 lần khởi tạo ngẫu nhiên tâm cụm khác nhau (luôn với cùng số lượng tâm cụm K) và thu được 50 bộ tâm cụm khác nhau. Đầu là cách được gọi ý cho việc chọn 1 kết quả từ 50 kết quả trên để sử dụng?

Select one:

- ☐ a. Chỉ có cách duy nhất để chọn là yêu cầu dữ liệu phải có nhãn  $y_i$
- ☒ b. Chọn kết quả mà  $\frac{1}{N} \sum_{i=1}^N \|x_i - m_{x_i}\|^2$  đạt giá trị nhỏ nhất trong 50 lần, với  $m_{x_i}$  là tâm cụm mà  $x_i$  được gán vào.
- ☐ c. Luôn chọn lần cuối cùng (thứ 50), vì lần này có khả năng đã hội tụ thành một giải pháp tốt

Time left 0:57:55

Question 8  
Not yet  
answered  
Marked out of  
1.00

Hãy chọn phát biểu đúng về Quá khớp (Overfitting)

Select one:

- ☐ a. Một hàm được gọi là bị quá khớp so với hàm khác nếu nó xấp xỉ kém chính xác hơn trên tập dữ liệu huấn luyện, nhưng chính xác hơn khi phân đoán các dữ liệu trong tương lai
- ☐ b. Một hàm được gọi là bị quá khớp so với hàm khác nếu nó xấp xỉ kém chính xác hơn trên tập dữ liệu huấn luyện và phân đoán các dữ liệu trong tương lai kém chính xác hơn
- ☒ c. Một hàm được gọi là bị quá khớp so với hàm khác nếu nó xấp xỉ chính xác hơn trên tập dữ liệu huấn luyện, nhưng kém chính xác hơn khi phân đoán các dữ liệu trong tương lai
- ☐ d. Tất cả các lựa chọn khác là sai

Question 9  
Not yet  
answered  
Marked out of  
1.00

Phương pháp bình phương tối thiểu chỉ có thể áp dụng được khi hàm hồi quy có độ lệch (slope/bias) dương?

Select one:

- ☐ True
- ☒ False

Question 10  
Not yet  
answered  
Marked out of  
1.00

Những phát biểu nào sau đây là SAI khi huấn luyện bộ phân loại SVM tuyến tính từ một tập huấn luyện kích thước  $m$ ?

Select one or more:

- ☒ a. Có thể không có vector hỗ trợ
- ☐ b. Số lượng vector hỗ trợ phụ thuộc vào chọn hằng số phạt  $C$
- ☒ c. Số lượng vector hỗ trợ phụ thuộc vào  $m$
- ☒ d. Bộ phân loại học được có thể đưa ra dự đoán sai cho tất cả tập dữ liệu huấn luyện
- ☐ e. Số lượng vector hỗ trợ độc lập với  $m$

Question 11  
Not yet  
answered  
Marked out of  
1.00

Với soft-margin SVM, những mẫu nào sẽ có giá trị  $\xi_i$  khác 0

Select one or more:

- ☐ a. Tất cả các mẫu nằm trên siêu phẳng lề
- ☐ b. Tất cả các mẫu nằm ngoài lề
- ☒ c. Tất cả các mẫu bị sai lớp
- ☒ d. Tất cả các mẫu nằm bên trong lề

Question 12  
Not yet  
answered  
Marked out of  
1.00

Trong hồi quy tuyến tính, ta thường muốn tìm một hàm  $f(x; \mathbf{w}) = w_0 + w_1x_1 + \dots + w_nx_n$  từ không gian hàm  $\mathbf{F}$ , trong đó  $\mathbf{w} = (w_0, w_1, \dots, w_n)$  thuộc một không gian tham số  $\mathbf{W}$ . Ta có thể nói gì về  $\mathbf{F}$  và  $\mathbf{W}$ ?

Select one:

- ☐ a. Chúng có thể giao nhau
- ☐ b. Chúng không giao nhau
- ☐ c. Chúng là một
- ☒ d. Ta không thể nói chúng giao nhau hay không giao nhau

Question 13  
Not yet  
answered  
Marked out of  
1.00

Mô hình hồi quy sau đây thuộc loại nào?  $f(x) = w_0 + w_1x_1 + \dots + w_nx_n$  trong đó  $w_0, w_1, \dots, w_n$  là các hệ số hồi quy

Select one:

- ☒ a. Mô hình tuyến tính
- ☐ b. Mô hình phi tuyến
- ☐ c. Mô hình phi tham số





Question 19  
Not yet  
answered  
Marked out of  
1.00

Phát biểu nào sau đây là đúng về giải thuật học láng giềng gần nhất?

Select one:

- ☐ a. Khi số lượng các thuộc tính kiểu số là ít, giải thuật học láng giềng gần nhất đạt độ chính xác cao hơn nếu các thuộc tính kiểu số đó có cùng khoảng giá trị.
- ☐ b. Bất kể số lượng các thuộc tính kiểu số, giải thuật học láng giềng gần nhất đạt độ chính xác cao hơn nếu các thuộc tính kiểu số đó có cùng khoảng giá trị. ✓
- ☐ c. Khi số lượng các thuộc tính kiểu số là nhiều, giải thuật học láng giềng gần nhất đạt độ chính xác cao hơn nếu các thuộc tính kiểu số đó có cùng khoảng giá trị.

Question 20  
Not yet  
answered  
Marked out of  
1.00

Phát biểu nào sau đây là SAI về Maximum A Posteriori estimation (MAP)?

Select one or more:

- ☐ a. MAP cần biết vài tri thức về các (siêu) tham số của mô hình. ✓
- ☐ b. MAP có thể được dùng để học một mô hình từ dữ liệu.
- ☐ c. MAP không cần biết bất kỳ tri thức nào về các (siêu) tham số của mô hình.
- ☐ d. MAP có thể ước lượng phân bố hậu nghiệm đầy đủ từ một tập dữ liệu.
- ☐ e. MAP có thể thực hiện suy diễn và phân đoán cho các ví dụ (quan sát) cụ thể. ✓

Question 21  
Not yet  
answered  
Marked out of  
1.00

Đánh giá mô hình là

Select one:

- ☐ a. việc đánh giá hiệu quả (chất lượng) của một mô hình hoặc phương pháp phân tích dữ liệu, chỉ bằng cách sử dụng các kích bản thực tế.
- ☐ b. việc đánh giá hiệu quả (chất lượng) của một mô hình hoặc phương pháp phân tích dữ liệu, bằng cách sử dụng một hoặc nhiều tập dữ liệu. ✓
- ☐ c. việc khám phá của chúng ta về một mô hình đã được học để tìm ra tri thức mới.

Question 22  
Not yet  
answered  
Marked out of  
1.00

Với một tập dữ liệu huấn luyện rất lớn, thuật toán nào có thời gian huấn luyện nhỏ nhất?

Select one:

- ☐ a. Random forests
- ☒ b. K-NN. ✓
- ☐ c. linear SVM
- ☐ d. Neural nets

Question 23  
Not yet  
answered  
Marked out of  
1.00

Hãy chọn phát biểu đúng về Kém khớp (Underfitting).

Select one:

- ☐ a. Một hàm được gọi là bị kém khớp so với hàm khác nếu nó xấp xỉ chính xác hơn trên tập dữ liệu huấn luyện, nhưng kém chính xác hơn khi phân đoán các dữ liệu trong tương lai.
- ☐ b. Một hàm được gọi là bị kém khớp so với hàm khác nếu nó xấp xỉ kém chính xác hơn trên tập dữ liệu huấn luyện và phân đoán các dữ liệu trong tương lai kém chính xác hơn. ✓
- ☐ c. Tất cả các lựa chọn khác là sai.
- ☐ d. Một hàm được gọi là bị kém khớp so với hàm khác nếu nó xấp xỉ kém chính xác hơn trên tập dữ liệu huấn luyện, nhưng chính xác hơn khi phân đoán các dữ liệu trong tương lai.

Question 24  
Not yet  
answered  
Marked out of  
1.00

Thuật toán Apriori làm những gì?

Select one or more:

- ☐ a. Sinh ra những luật có confidence lớn hơn ngưỡng minconf. ✓
- ☐ b. Sinh ra những luật có confidence nhỏ hơn ngưỡng minconf.
- ☐ c. Cắt đi những luật có support lớn hơn ngưỡng minsup.
- ☐ d. Cắt đi những luật có support nhỏ hơn ngưỡng minsup. ✓

Question 25  
Not yet  
answered  
Marked out of  
1.00

Cực đại hoá khả năng (Maximum Likelihood estimation) có thể được sử dụng để

Select one or more:

- ☐ a. thực hiện suy diễn hoặc phân đoán cho một ví dụ (quan sát).
- ☐ b. ước lượng khả năng cực đại của một mô hình.
- ☐ c. suy diễn sự chính xác của một mô hình. ✓
- ☐ d. học một mô hình từ một tập huấn luyện.
- ☐ e. khám phá tri thức của một mô hình đã học. ✓

Question 26  
Not yet  
answered  
Marked out of  
1.00

Phát biểu nào sau đây là đúng về hàm khoảng cách Manhattan?

Select one:

- ☐ a. Hàm khoảng cách Manhattan được dùng cho các biến kiểu định danh
- ☐ b. Hàm khoảng cách Manhattan được dùng cho các biến (thuộc tính) kiểu liên tục
- ☐ c. Hàm khoảng cách Manhattan không dùng được cho các biến kiểu liên tục hoặc kiểu định danh
- ☒ d. Hàm khoảng cách Manhattan được dùng cho các biến kiểu liên tục hoặc kiểu định danh

?

?

Question 27  
Not yet  
answered  
Marked out of  
1.00

Một cụm là

Select one:

- ☐ a. phép toán làm việc trên cơ sở dữ liệu để biến đổi đơn giản hóa dữ liệu nhằm chuẩn bị cho thuật toán máy học hoặc khai phá dữ liệu
- ☐ b. là nhóm các đối tượng giống và khác nhau
- ☐ c. biểu diễn tương trưng về các sự kiện hoặc ý tưởng mà từ đó thông tin có thể được trích xuất
- ☐ d. là một nhóm các đối tượng giống nhau và chúng khác biệt lớn với các đối tượng của nhóm khác

X

X

X

Question 28  
Not yet  
answered  
Marked out of  
1.00

Với một cây quyết định học được, thì nhân phân loại đối với 1 ví dụ mới được xác định bởi việc duyệt (đi theo) ...

Select one:

- ☐ a. một đường đi duy nhất trong cây
- ☐ b. đường đi có độ sâu lớn nhất trong cây
- ☐ c. tất cả các đường đi trong cây
- ☐ d. đường đi có độ sâu nhỏ nhất trong cây

✓

Question 29  
Not yet  
answered  
Marked out of  
1.00

Giả sử rằng chúng ta có một cơ sở dữ liệu trong đó mỗi giao dịch chứa cả tên và giá của từng item. Thông tin về giá có ảnh hưởng đến chất lượng của thuật toán Apriori không?

Select one:

- ☐ True
- ☒ False

Question 30  
Not yet  
answered  
Marked out of  
1.00

Tại sao bước tìm tập mục thường xuyên với kích thước 2 (2-itemsets) thường là khó nhất?

Select one:

- ☐ a. Đây luôn là bước bắt buộc để tìm bất kỳ tập mục thường xuyên lớn hơn.
- ☒ b. Với một xác suất cao, nó yêu cầu bộ nhớ lớn nhất so với các loại tập mục thường xuyên kích thước khác.
- ☐ c. Lý do chính là thời gian tính toán lớn

Question 31  
Not yet  
answered  
Marked out of  
1.00

Hàm đánh giá Information Gain có xu hướng ...

Select one or more:

- ☐ a. đẩy các thuộc tính có độ giảm ít hơn về mức độ hỗn tạp về phía gần hơn các nút lá
- ☐ b. đẩy các thuộc tính có độ giảm nhiều hơn về mức độ hỗn tạp về phía gần hơn nút gốc
- ☐ c. đẩy các thuộc tính có độ giảm ít hơn về mức độ hỗn tạp về phía gần hơn nút gốc
- ☐ d. đẩy các thuộc tính có độ giảm nhiều hơn về mức độ hỗn tạp về phía gần hơn các nút lá

✓

✓

Question 32  
Not yet  
answered  
Marked out of  
1.00

Rời rạc hoá dữ liệu (Data discretization) là

Select one:

- ☐ a. chuyển đổi từ liên tục thành rời rạc
- ☐ b. tìm loại lai (outliers)
- ☐ c. chuyển đổi dữ liệu từ rời rạc thành liên tục
- ☐ d. mở rộng dữ liệu (scale up data)

✓

Question 33  
Not yet answered  
Marked out of 1.00

Cho trước các tập mục kích thước là 2 (2-itemsets) và tổng số hỗ trợ (support count) của chúng như trong bảng sau đây.

2-itemset	Count
{11, 12}	6
{11, 13}	7
{11, 14}	4
{12, 13}	6
{12, 14}	9
{13, 14}	6

Nếu ngưỡng của support count là 6, theo thuật toán Apriori, tập mục 3-itemset nào sau đây không phải là tập mục thường xuyên?

Select one:

- ☐ a. {12, 13, 14}
- ☒ b. {11, 12, 14}
- ☐ c. tất cả các tập mục kích thước 3 (3-itemsets)
- ☐ d. {11, 12, 13}

Previous page

Time left 0:56:33

Question 34  
Not yet answered  
Marked out of 1.00

Phương pháp bình phương tối thiểu học một hàm  $f(x) = w_0 + w_1x_1 + \dots + w_nx_n$  từ một tập học có cỡ  $M$  bằng cách tìm vectơ  $\mathbf{w}^* = (w_0^*, w_1^*, \dots, w_n^*)$ , trong đó

Select one:

- ☐ a.  $\mathbf{w}^* = \arg \min_{\mathbf{w}} \sum_{i=1}^M (y_i - w_0 - w_1x_{i1} - \dots - w_nx_{in})^2 + \lambda \|\mathbf{w}\|_2^2$
- ☐ b.  $\mathbf{w}^* = \arg \min_{\mathbf{w}} \sum_{i=1}^M (y_i - w_0 - w_1x_{i1} - \dots - w_nx_{in})$
- ☒ c.  $\mathbf{w}^* = \arg \min_{\mathbf{w}} \sum_{i=1}^M (y_i - w_0 - w_1x_{i1} - \dots - w_nx_{in})^2$

Time left 0:56:30

Question 35  
Not yet answered  
Marked out of 1.00

Đối với giải thuật học láng giềng gần nhất, điều gì xảy ra nếu sử dụng một số lượng rất lớn các láng giềng gần nhất để đưa ra dự đoán?

Select one or more:

- ☐ a. Kết quả dự đoán sẽ có xu hướng trở nên chính xác hơn
- ☒ b. Kết quả dự đoán sẽ có xu hướng là hiển nhiên (trivial)
- ☐ c. Giải thuật học láng giềng gần nhất sẽ có xu hướng học quá khớp (overfitting)
- ☒ d. Giải thuật học láng giềng gần nhất sẽ có xu hướng học chưa khớp (underfitting)

Question 36  
Not yet answered  
Marked out of 1.00

Một mạng nơ-ron nhân tạo có thể được coi là một hàm phi tuyến. Hàm phi tuyến đó theo gì?

Select one:

- ☐ a. Cả tín hiệu vào  $x$  và bộ trọng số
- ☐ b. Bộ trọng số
- ☒ c. Tín hiệu đầu vào  $x$

Question 37  
Not yet answered  
Marked out of 1.00

Nhắc lại hàm mục tiêu của soft-margin SVM có dạng  $\|\mathbf{w}\|^2 + C \sum_i \xi_i$ . Giảm giá trị của  $C$  có thể dẫn đến

Select one:

- ☐ a. tăng khả năng overfitting
- ☒ b. giảm khả năng overfitting
- ☐ c. ít nhạy cảm với ngoại lai (outliers)
- ☒ d. nhỏ hơn

Question 38  
Not yet answered  
Marked out of 1.00

Thuật toán K-means

Select one or more:

- ☒ a. có thể hội tụ tới các cụm khác nhau, phụ thuộc vào cách khởi tạo tâm cụm
- ☐ b. luôn hội tụ tới một cách phân cụm duy nhất
- ☒ c. nhạy cảm với dữ liệu ngoại lai (outliers)



Question 39  
Not yet  
answered  
Marked out of  
1.00

Tất cả các mạng nơ-ron tính toán các hàm phi tuyến theo bộ tham số của nó?

Select one:

- ☐ True  
☒ False

?

Question 40  
Not yet  
answered  
Marked out of  
1.00

Giảm chiều dữ liệu có thể được sử dụng như một cách tiền xử lý cho các thuật toán học máy như cây quyết định, K-means, mạng nơ-ron, etc.?

Select one:

- ☐ True  
☒ False

Question 41  
Not yet  
answered  
Marked out of  
1.00

Những phát biểu nào sau đây là đúng về bài toán đối ngẫu và SVM?

Select one or more:

- ☐ a. Chúng ta sử dụng nhân tử Lagrange trong vấn đề tối ưu với ràng buộc bất đẳng thức tuyến tính  
☒ b. Khi giải quyết bài toán đối ngẫu của SVM, chúng ta chỉ cần tích vô hướng của  $x_i$  và  $x_j$  cho tất cả  $i, j$ , và không cần thông tin của  $x_i$ .  
☐ c. Chúng ta sử dụng nhân tử Lagrange trong vấn đề tối ưu với ràng buộc bất đẳng thức bậc 2

neu giai quyet la prediction thi b dung

Question 42  
Not yet  
answered  
Marked out of  
1.00

Phân loại Bayes là

Select one:

- ☐ a. Một lớp các thuật toán mà sử dụng tính toán về tần suất xuất hiện của các giá trị thuộc tính trong tập huấn luyện  
☐ b. Một lớp các thuật toán mà sử dụng các tính toán về xác suất hậu nghiệm  
☒ c. Một lớp các thuật toán mà sử dụng information gain

Question 43  
Not yet  
answered  
Marked out of  
1.00

Học máy (Machine Learning) cung cấp các phương pháp để phân tích dữ liệu, tạo các phân đoán cho các quan sát trong tương lai

Select one:

- ☐ a. Đúng  
☐ b. Sai, nó cung cấp các phương pháp tăng tốc tính toán của máy tính  
☒ c. Đúng, nó còn cung cấp các phương pháp tăng tốc tính toán của máy tính

?

Question 44  
Not yet  
answered  
Marked out of  
1.00

Phát biểu nào sau đây là phù hợp nhất về lựa chọn mô hình (model selection)?

Select one:

- ☐ a. Lựa chọn mô hình quan tâm đến việc tìm thiết đặt tốt nhất về bộ (siêu) tham số trong một mô hình khi huấn luyện nó từ một tập dữ liệu. Đôi khi nó cũng nói đến việc lựa chọn một trong số các mô hình đang có.  
☐ b. Các phát biểu khác đều sai.  
☒ c. Lựa chọn mô hình chỉ quan tâm đến việc lựa chọn một mô hình tốt nhất từ một tập đang có.

Question 45

Not yet  
answered  
Marked out of  
1.00

Hãy chọn các phát biểu đúng

Select one or more:

- ☐ a. Đánh giá chéo có thể được dùng để đánh giá hiệu quả và lựa chọn tham số cho một mô hình học máy  
☐ b. Đánh giá chéo không tốn kém thời gian  
☐ c. Đánh giá chéo có thể tốn thời gian  
☐ d. Đánh giá chéo không bị ảnh hưởng xấu từ sự ngẫu nhiên, nghĩa là việc chia tập dữ liệu một cách ngẫu nhiên không làm ảnh hưởng nhiều đến kết quả đánh giá  
☒ e. Đánh giá chéo (Cross validation) chỉ được sử dụng để đánh giá hiệu quả của một mô hình học máy