

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
$\{i_1, i_2\}$	6
$\{i_1, i_3\}$	7
$\{i_1, i_4\}$	4
$\{i_2, i_3\}$	6
$\{i_2, i_4\}$	9
$\{i_3, i_4\}$	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.  $\{i_1, i_2, i_3\}$
- ☐ b.  $\{i_2, i_3, i_4\}$
- ☐ c. tất cả các tập mục kích thước 3 (3-itemsets)
- ☐ d.  $\{i_1, i_2, i_4\}$

# IT3190 - Nhập môn Học máy và khai phá dữ liệu

Quiz navigation

TRUNG TRINH QUANG

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45			

Question 3

Not yet answered

Marked out of 1.00

Một nút thắt cổ chai nghiêm trọng của thuật toán Apriori là

Select one:

- ☐ a. sinh ứng viên
- ☐ b. số vòng lặp
- ☐ c. tìm các tập mục thường xuyên
- ☐ d. cắt tỉa

Previous page

Time left 0:59

Next page



# máy và khai phá dữ liệu

Time left 0:49:53

Hãy chọn các phát biểu đúng

Select one or more:

- ☐ a. Đánh giá chéo (Cross validation) chỉ được sử dụng để đánh giá hiệu quả của một mô hình học máy
- ☐ b. 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
- ☐ c. Đánh giá chéo không tốn kém thời gian
- ☐ d. Đánh giá chéo có thể được dùng để đánh giá hiệu quả và lựa chọn tham số cho một mô hình học máy
- ☐ e. Đánh giá chéo có thể tốn thời gian

Next page

## c máy và khai phá dữ liệu

Time left 0:47:48

41

ed  
out of

Dưới đây là một số lựa chọn để có thể thực hiện khi huấn luyện mạng nơ-ron. Chọn tất cả các trường hợp sẽ khiến mạng của bạn KHÔNG đạt được độ chính xác cao trong tương lai.

Select one or more:

- ☐ a. Đảo ngẫu nhiên lại dữ liệu khi bắt đầu mỗi epoch
- ☐ b. Chuẩn hóa dữ liệu huấn luyện nhưng không thay đổi dữ liệu kiểm thử
- ☐ c. Khởi tạo tất cả bộ tham số bằng 0
- ☐ d. Sử dụng momentum

page

Next page



# IT3190 - Nhập môn Học máy và khai phá dữ liệu

Quiz navigation

TRUNG TRINH QUANG

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45			

Not yet answered

Question 5

Not yet answered

Marked out of 1.00

Phát biểu nào sau đây là đúng về hồi quy Ridge và Lasso?

Select one:

- ☐ a. Hồi quy Ridge có thể giúp loại bỏ một tập con các thuộc tính không quan trọng
- ☐ b. Cả Ridge và Lasso đều có thể giúp loại bỏ một tập con các thuộc tính không quan trọng
- ☐ c. Lasso có thể giúp loại bỏ một tập con các thuộc tính không quan trọng

Previous page

Time left 0:59:08

Next page

Time left 0:48:28

37

Trong khai phá luật kết hợp, tổng số hỗ trợ (support count) của tập mục (itemset) A được định nghĩa là gì?

Select one:

- ☐ a. Tỷ lệ tổng số giao dịch không chứa A và tổng số tất cả giao dịch
- ☐ b. Tổng số lượng giao dịch chứa A
- ☐ c. Tổng số giao dịch không chứa A
- ☐ d. Tỷ lệ tổng số giao dịch chứa A và tổng số tất cả giao dịch

Next page



# ọc máy và khai phá dữ liệu

Time left 0:5

Question 11

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. Giải thuật học láng giềng gần nhất sẽ có xu hướng học quá khớp (overfitting)
- ☐ 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 chưa khớp (underfitting)
- ☐ d. Kết quả dự đoán sẽ có xu hướng trở nên chính xác hơn

Previous page

Next page

Time left 0:47:53

40

d  
out of

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

Select one:

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

s page

Next page



Time left 0:49:38

23

d  
cut of

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 bậc 2
- ☐ b. Chúng ta sử dụng nhân tử Lagrange trong vấn đề tối ưu với ràng buộc bất đẳng thức tuyến tính
- ☐ c. 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  $x_i$ .

page

Next page

## Thôn Học máy và khai phá dữ liệu

Time left

Question 31

Not yet  
answered

Marked out of  
1.00

Những khía cạnh nào sau đây là các ưu điểm của cây quyết định?

Select one or more:

- ☐ a. Không giờ gặp phải vấn đề học quá khớp
- ☐ b. Có thể làm việc rất tốt đối với các bài toán phân loại trong đó các ví dụ được biểu diễn bởi rất nhiều các thuộc tính
- ☐ c. Dễ dàng giải thích cho các dự đoán được đưa ra
- ☐ d. Có thể sử dụng được với 1 tập huấn luyện chứa các ví dụ thiếu giá trị
- ☐ e. Một giải thuật học rất ổn định (stable)

Previous page

Next page



Time left 0:48:22

Question 35

Not yet  
answered

Marked out of  
1.00

Chúng ta muốn phân 7 điểm dữ liệu vào trong 3 cụm sử dụng thuật toán K-means (với khoảng cách Euclid). Giả sử rằng sau vòng lặp đầu tiên, các cụm C1, C2 và C3 chứa các điểm dữ liệu sau (trong không gian 2 chiều): C1 chứa 2 điểm dữ liệu:  $\{(0,6), (6,0)\}$ ; C2 chứa 3 điểm dữ liệu:  $\{(2,2), (4,4), (6,6)\}$ ; C3 chứa 2 điểm dữ liệu:  $\{(5,5), (7,7)\}$ .

Tâm của 3 cụm sẽ là?

Select one:

- ☐ a. C1: (6,6), C2: (12,12), C3: (12,12)
- ☐ b. C1: (0,0), C2: (48,48), C3: (35,35)
- ☐ c. C1: (3,3), C2: (4,4), C3: (6,6)
- ☐ d. C1: (3,3), C2: (6,6), C3: (12,12)

Previous page

Next page



ENG

7:04 PM  
2/14/2022



Question 27

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

Previous page

Next page



# Học máy và khai phá dữ liệu

Time left 0:58:06

Question 13

Not yet  
answered

Marked out of  
1.00

Giá trị trả về của hàm khoảng cách Euclidean giữa 2 điểm dữ liệu  $X(1,7)$  và  $Y(4,3)$  là bao nhiêu?

Select one:

- ☐ a. 1
- ☐ b. 5
- ☐ c. 3
- ☐ d. 9
- ☐ e. 7

Previous page

Next page

# Học máy và khai phá dữ liệu

Time left 0:49:2

Question 25

Not yet  
answered

Marked out of  
1.00

Những phát biểu nào sau đây SAI về thuật toán K-means

Select one or more:

- ☐ a. K-means không thể bị quá khớp (overfitting) do không có hàm lỗi nào được sử dụng cho huấn luyện
- ☐ b. K-means có thể chịu kết quả học tối (underfitting) khi tăng số lượng cụm
- ☐ c. Khi số lượng cụm tăng, lỗi huấn luyện sẽ có thể giảm
- ☐ d. Hiệu năng của thuật toán K-means có thể thay đổi khi chạy các lần khác nhau với các khởi tạo tâm cụm khác nhau
- ☐ e. Hiệu năng của thuật toán K-means có thể phụ thuộc vào chọn độ đo khoảng cách giữa 2 điểm

Previous page

Next page



Question 26

Not yet  
answered  
Marked out of  
1.00

Phân loại Bayes là

Select one:

- ☐ a. Một lớp các thuật toán mà sử dụng các tính toán về xác suất hậu nghiệm
- ☐ b. Một lớp các thuật toán mà sử dụng information gain
- ☐ c. 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

Previous page

Next page

Question 29

not yet  
answered  
marked out of  
1.0

Thuật toán gradient descent làm gì?

Select one:

- ☐ a. Tính tích lũy giá trị lỗi trong các epochs
- ☐ b. Tìm kiến trúc thích hợp nhất với một bài toán cho trước
- ☐ c. Tính tự động số lượng epochs cần thiết để huấn luyện mạng nơ-ron
- ☐ d. Tìm tham số mô hình mà cực tiểu hàm lỗi

Previous page

Next page



Question 10

Not yet  
answered

Marked out of  
1.00

Hàm kích hoạt là hàm gì?

Select one:

- ☐ a. Tính tổng có trọng số của các tín hiệu đầu vào của một nơ-ron
- ☐ b. Là hàm tính đầu ra cho mạng nơ-ron
- ☐ c. Là hàm tính đầu ra của một nơ-ron khi cho trước đầu vào tổng hợp (Net) của nơ-ron đó
- ☐ d. Là hàm tính đầu ra cho một nơ-ron khi cho trước tín hiệu đầu vào của nơ-ron đó

Previous page

## IT3190 - Nhập môn Học máy và khai phá dữ liệu

Quiz navigation



TRUNG TRINH QUANG

1	2	3	4	5	6
7	8	9	10	11	
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45			

Not yet answered

Question 4

Not yet answered

Marked out of 1.00

Phát biểu nào sau đây sai?

Select one:

- ☐ a. Tiền xử lý thường được yêu cầu trước bất kỳ một bước phân tích sâu hơn trên một tập dữ liệu cho trước
- ☐ b. Tiền xử lý dữ liệu toàn bộ thông tin từ tập dữ liệu gốc
- ☐ c. Tiền xử lý thường dẫn đến một số mất mát thông tin trên một tập dữ liệu cho trước

Previous page

Next page

Finish attempt...



Question 8

Not yet  
answered

Marked out of  
1.00

Những phát biểu nào SAI về độ phức tạp của thuật toán phân cụm K-means?

Select one or more:

- ☐ a. Nó phụ thuộc nghiêm trọng vào độ đo khoảng cách
- ☐ b. Nó độc lập với số chiều của dữ liệu
- ☐ c. Nó là tuyến tính theo số chiều và kích thước tập huấn luyện
- ☐ d. Thời gian tính toán phụ thuộc lớn vào số lượng cụm
- ☐ e. Nó có thể yêu cầu một lượng lớn số lần lặp

Previous page

## lọc máy và khai phá dữ liệu

Time left 0:47:33

Question 44

not yet  
answered

marked out of  
10

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

Previous page

Next page



máy và khai phá dữ liệu

Time left 0:49:59

Giả sử có ma trận nhầm lẫn dưới đây. Độ chính xác của bộ phân loại là bao nhiêu nếu nó phán đoán một người có mua máy tính hay không?

	<i>Buy_computer = yes</i>	<i>Buy_computer = no</i>	<i>Total</i>
<i>Buy_computer = yes</i>	6954	46	7000
<i>Buy_computer = no</i>	412	2588	3000
<i>Total</i>	7366	2634	10000

Select one:

- ☐ a.  $(46 + 412)/10000$
- ☐ b.  $(6954 + 2588)/10000$
- ☐ c.  $(6954 + 46)/10000$
- ☐ d.  $(6954 + 412)/10000$

Next page



## Question 33

Not yet  
answeredMarked out of  
1.00

Xét tập học sau đây, trong đó  $x_1$  và  $x_2$  là các biến (thuộc tính) đầu vào và  $c_1$  và  $c_2$  là các nhãn lớp.

$x_1$	$x_2$	Class
-1	1	$c_2$
0	1	$c_1$
0	2	$c_2$
1	-1	$c_2$
1	0	$c_1$
1	2	$c_1$
2	2	$c_2$
2	3	$c_1$

Áp dụng giải thuật học láng giềng gần nhất và hàm khoảng cách Euclid và sử dụng 3 láng giềng gần nhất, hãy xác định nhãn lớp cho ví dụ ( $z_1 = 1, z_2 = 1$ )?

Select one:

- ☐ a.  $c_1$   
☐ b.  $c_2$   
☐ c. Không xác định



## Nhập môn Học máy và khai phá dữ liệu

Question 7

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 có thể quan sát được trong thực tế
- ☐ c. một sự kiện mà có thể quan sát được hoặc không
- ☐ d. một sự kiện mà không thể quan sát được

Previous page

## Học máy và khai phá dữ liệu

Time left 0:47:4

Question 43

Answer saved

Marked out of  
1.00

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

Select one:

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

Clear my choice

Previous page

Next page



## ọc máy và khai phá dữ liệu

Time left 0:49:45

Question 22

not yet  
answered  
marked out of  
0

Những cách nào sau đây được xem là giải pháp cho khuyết dữ liệu (missing value)?

Select one or more:

- ☐ a. Điền thủ công những ô thiếu dữ liệu
- ☐ b. Sử dụng giá trị trung bình (mean) và trung vị (median) cho các điểm thuộc cùng lớp
- ☐ c. Chuẩn hóa dữ liệu (Normalization)

Previous page

Next page

## c máy và khai phá dữ liệu

Time left 0:47:57

39

Chuỗi hành động nào sau đây là phù hợp nhất khi bạn muốn khám phá tri thức mới từ dữ liệu?

Select one:

- ☐ a. Mô hình hoá, thu thập dữ liệu, tiền xử lý dữ liệu, đánh giá kết quả
- ☐ b. Thu thập dữ liệu, tiền xử lý dữ liệu, đánh giá kết quả, mô hình hoá, phân tích
- ☐ c. Thu thập dữ liệu, tiền xử lý dữ liệu, mô hình hoá, đánh giá kết quả

s page

Next page



Time left 0:48:30

34

Khi mô hình hoá dữ liệu, sự không chắc chắn có thể xuất hiện trong

Select one:

- ☐ a. tập dữ liệu, bộ tham số của một mô hình (hàm), và lớp mô hình đang dùng
- ☐ b. tập dữ liệu thu thập được
- ☐ c. lớp mô hình (lớp hàm) đang dùng
- ☐ d. bộ tham số của một mô hình (hàm) đã học được

us page

Next page

## c máy và khai phá dữ liệu

Time left 0:43:45

32

ut of

Trong quá trình phát triển cây quyết định cho bài toán phân loại, độ phân biệt (discriminateness) được dùng để lựa chọn thuộc tính kiểm tra cho 1 nút trung gian. Một thuộc tính có độ phân biệt cao ...

Select one:

- ☐ a. nhằm tăng độ hỗn tạp của tập học khi được phân chia vào các nhánh mới.
- ☐ b. nhằm giảm độ hỗn tạp của tập học khi được phân chia vào các nhánh mới.
- ☐ c. nhằm giảm độ phân biệt (discriminateness) của tập học khi được phân chia vào các nhánh mới.

page

Next page