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 là hiển nhiên (trivial)

b. Giải thuật học láng giềng gần nhất sẽ có xu hướng học quá khớp (overfitting)

c. Giải thuật học láng giêng gần nhất sẽ có xu hướng học chưa khớp (underfitting)



d. Kết quả dự đoán sẽ có xu hướng trở nên chính xác hơn

EXAMINATION HUST

IT3190 - Nhập môn Học máy và khai phá dữ liệu





Not yet answered

Marked out of 1.00

Previous page

Trong hồi quy tuyến tính, ta thường muốn tìm một hàm $f(x; \mathbf{w}) = w_0 + w_1 x_1 + \dots + w_n x_n$ 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:

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

\-, \tau \W=\W.

f(x1) = 47.4 b

1

F: Ko gian han tryan tich N " Ko gran rector tham 50

Time left 0:42:43

Question 4

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:

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

Previous page

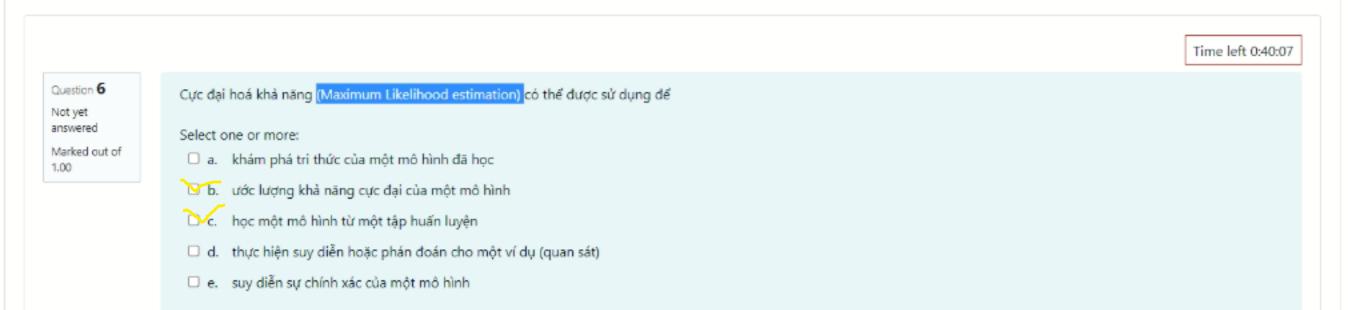
Not yet answered

Marked out of 1.00 Hãy chọn phát biểu đúng về Kém khớp (Underfitting)

Select one:

- O a. Một hàm được gọi là bị kém khớp so với hàm khác nếu nó xấp xì kém chính xác hơn trên tập dữ liệu huấn luyện và phán đoán các dữ liệu trong tương lai kém chính xác hơn
- 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, nhưng chính xác hơn khi phán đoán các dữ liệu trong tương lai
- O 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 xí chính xác hơn trên tập dữ liệu huấn luyện, nhưng kém chính xác hơn khi phán đoán các dữ liệu trong tương lai

Previous page



Previous page



Not yet answered

Marked out of 1.00

Những phát biểu nào sau đây đúng về SVM?

Select one or more:



🗹 a. Tāng tham số C hướng tới giảm nhạy cảm với (outliers)



√б. Tăng tham số С hướng tới giảm lỗi huấn luyện



igtharpoonup Hard-margin SVM là một trường hợp đặc biệt của soft-margin SVM khi C được cho bằng 0



 $oxed{\mathbb{Z}}$ d. Tăng tham số C hướng tới giảm lề

Previous page

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. K-NN

O b. Random forests

O c. Neural nets

O d. linear SVM

Previous page



Previous page



Marked out of 1.00

Not yet

Đâu không phải là phương pháp làm sạch dữ liệu?

Select one:

- O a. Phân cụm (Clustering)
- O b. Hồi quy (Regression)
- c. Phân khoảng (Binning)



Previous page



Not yet answered Marked ou

Marked out of 1.00 Mô hình hồi quy sau đây thuộc lớp nào? $f(x)=w_0+w_1x_1+\cdots+w_nx_n$ trong đó w_0,w_1,\ldots,w_n là các hệ số hồi quy

Select one:

a. Mô hình tuyến tính

- O b. Mô hình phi tuyến
- c. Mô hình phi tham số

Previous page

Not yet answered

Marked out of 1.00 Khai phá tri thức là ...

Select one:

- O a. quá trình gom nhóm các đối tượng sao cho các đối tượng trong cùng một nhóm sẽ giống nhau hơn khi so với đối tượng thuộc nhóm khác
- O b. quá trình chuyển đổi dữ liệu từ dạng này sang một dạng cấu trúc khác
- O c. quá trình tìm ra những điểm dữ liệu mà thiếu giá trị trong một số trường thuộc tính
- ✓d. quá trình tạo ra tri thức từ một tập dữ liệu

Previous page



Time left 0:33:57

Question 13

Not yet answered

Marked out of 1.00

Số vòng lặp của thuật toán Apriori

Select one:

a. tăng theo kích thước của tập mục thường xuyên lớn nhất

- O b. tăng theo kích thước của dữ liệu
- c. giảm khi kích thước tập mục thường xuyên lớn nhất tăng kích thước

Clear my choice

Previous page

Time left 0:33:07

Question 14

Not yet answered

Marked out of 1.00 Trong khai phá luật kết hợp, tổng số hỗ trợ (support count) của tập mục (itemset) A được định nghĩa là gì?

Select one:

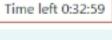
- O a. Tổng số giao dịch không chứa A
- O b. Tỷ lệ tổng số giao dịch không chứa A và tổng số tất cả giao dịch



O d. Tỷ lệ tổng số giao dịch chứa A và tổng số tất cả giao dịch

Clear my choice

Previous page



Not yet answered

Marked out of 1.00

Select one:

O a. Lý do chính là thời gian tính toán lớn

O b. Đây luôn là bước bắt buộc để tìm bất kỳ tập mục thường xuyên lớn hơn.

√c. Với một xác suất cao, nó yêu cầu bộ nhớ lớn nhất so với các loại tập thường xuyên kích thước khác.

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

Previous page

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:

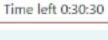
a. Khi số lượng các thuộc tính kiểu số là ít, giải thuật học láng giêng gần nhất đạt độ chính xác cao hơn nếu các thuộc tính kiểu số đó có cùng khoảng giá trị.



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.

O c. Khi số lượng các thuộc tính kiểu số là nhiều, giải thuật học láng giêng gần nhất đạt độ chính xác cao hơn nếu các thuộc tính kiểu số đó có cùng khoảng giá trị.

Previous page



Not yet answered Marked out

Marked out of 1.00 Kỹ thuật nào dưới đây có thể giúp giảm Overfitting?

Select one:

- O a. Bỏ bớt dữ liệu nếu nó có kích cỡ quá lớn
- O b. Thay đổi mô hình mới hoặc phương pháp học mới



Previous page



Not yet answered

Marked out of 1.00 Xác suất hậu nghiệm (Posterior probability) nói đến

Select one:

- Select one
- a. xác suất để một quan sát (sự kiện) xuất hiện khi biết trước một mô hình hoặc giả thuyết
- b. xác suất của một mô hình hoặc giả thuyết khi cho trước (một tập) dữ liệu
- O c. tri thức của một mô hình

Clear my choice

Next page

Previous page



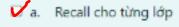
Time left 0:29:12

Question 20

Not yet answered

Marked out of 1.00 Công việc của bạn yêu cầu xây dựng một bộ phân loại mà có thể phán đoán một email cho trước là thường hay rác. Tuy nhiên bạn chỉ có thể thu thập được một tập dữ liệu rất mất cân đối, trong đó có 99.9% là email rác. Độ đo nào sau đây bạn nên dùng để đánh giá hiệu quả của bộ phân loại?

Select one or more:

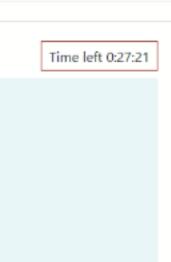




c. Trung bình bình phương lỗi (Mean squared error)

□ d. Accuracy

Previous page



Not yet answered

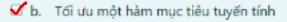
1.00

Marked out of

Những phát biểu nào sau đây là đúng về mạng nơ-ron?

Select one or more:

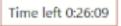
🗆 a. Có thể được thực hiện để hoạt động tốt ngay cả khi số lượng tham số lớn hơn nhiều so với số lượng điểm dữ liệu



c. Có thể sử dụng phối hợp các hàm kích hoạt khác nhau

☐ d. Chỉ được huấn luyện với thuật toán SGD (stochastic gradient descent)

Previous page



Not yet answered

Marked out of 1.00

Phát biểu nào sau đây là ĐÚNG về hạn chế của thuật toán phân loại (multinomial) Naive Bayes?

Select one or more:

a. Naive Bayes khó làm việc với những thuộc tính liên tục



b. Naive Bayes buộc phải tính xác xuất đầy đủ khi muốn đưa ra phán đoán nhân lớp cho một quan sát cụ thể

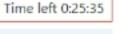


C. Naive Bayes giả thuyết các thuộc tính là độc lập với nhau, với điều kiện nhãn lớp cho trước



☐ d. Naive Bayes khó làm việc với các thuộc tính định danh mà có tập giá trị lớn

Previous page



Not yet answered Marked out of

1.00

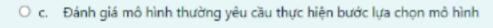
Phát biểu nào sau đây là SAI?

Select one:

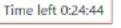
O a. Lựa chọn mô hình là một bước bắt buộc khi muốn so sánh nhiều mô hình (hoặc phương pháp) học máy khác nhau



Job. Đánh giá mô hình và lựa chọn mô hình trong Học máy là hai thứ độc lập với nhau



Previous page



Not yet answered

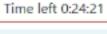
Marked out of 1.00 Tiền xử lý dữ liệu có thể xem như

Select one:

a. là quá trình biến đổi dữ liêu thô tới một dạng thích hợp để cho bước phân tích

- O b. phân tích một số tích chất chính của tập dữ liệu
- c. là quá trình quyết định định dạng về dữ liệu và thu thập dữ liệu

Previous page



Question 25 Not yet

answered

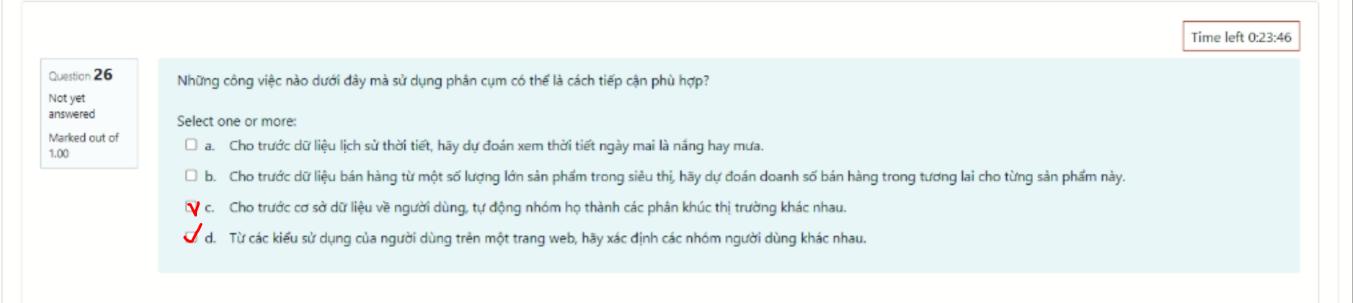
Marked out of 1.00

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

Select one: a. sinh ứng viên

- b. số vòng lặp
- O c. cắt tỉa
- d. tìm các tập mục thường xuyên

Previous page



Previous page



Not yet answered

Marked out of 1.00 Đặt S_1 và S_2 là tập các vectors hỗ trợ và w_1 và w_2 là các vector tham số học được cho bài toán phân tách tuyến tính sử dụng hard-margin và soft-margin SVM tương ứng. Những phát biểu nào sau đây đúng?

Select one or more:

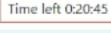
igvee a. w_1 có thể không bằng w_2 .

 \square b. S_1 có thể không là con của S_2

lacksquare c. $S_1\subset S_2$

 \square d. $w_1 = w_2$

Previous page



Not yet answered

Marked out of 1.00

Khi huấn luyện một mạng nơ-ron, chúng ta đang cố gắng tìm kiếm những gì?

Select one or more:

🗸 a. Các siêu tham số mà đưa ra hiệu năng tốt nhất



b. Bộ tham số (weights) tốt nhất trong không gian tham số



c. Một ước lượng của lỗi kiểm thừ



√d. Một hàm trong họ hàm mà có thể ánh xạ từ một ví dụ tới một đầu ra



e. Một thiết lập về kiến trúc mạng

Previous page

Time left 0:19:42

Question 29

Not yet answered

Marked out of 1.00

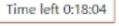
Những phát biểu nào sau đây SAI về thuật toán K-means

Select one or more:

- a. Khi số lượng cụm tăng, lỗi h ấn luyện sẽ có thể giảm
- □ b. 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 c. Hiệu năng của thuật toán K-means có thể phụ thuộc vào chọn độ đo khoảng cách giữa 2 điểm
- □ d. K-means có thể chịu kết quả học tồi (under tting) khi tăng số lượng cụm

e. K-means không thể bị quá khớp (overfitting) do không có hàm lỗi nào được sử dụng cho huấn luyện

Previous page



answered Marked out of 1.00

Not yet

và cùng số lượng đầu vào?

Một mạng truyền tiến nhiều tầng (multiple-layer neural network) với các hàm kích hoạt tuyến tính là tương đương với một mạng Perceptron mà sử dụng cùng hàm lỗi trên đầu ra

Select one:

O False

Previous page

Not yet answered Marked out o

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:

- O a. Information Gain
- ✓ b. Gain Ratio
- O c. Entropy

Previous page



Not yet answered

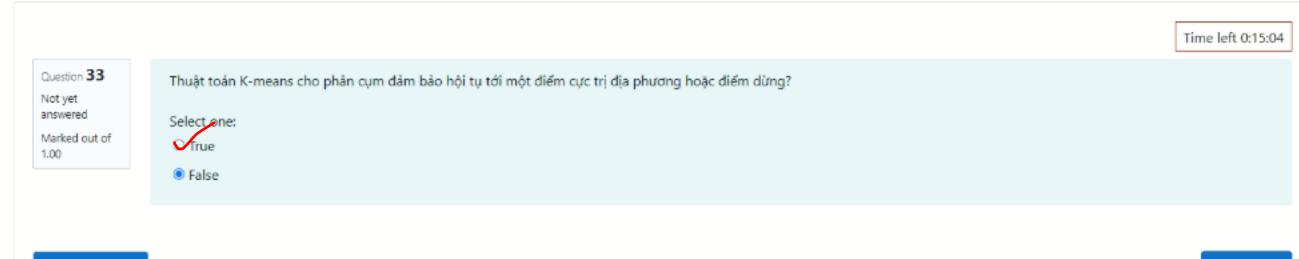
Marked out of 1.00 Trong hard-margin SVM cho dữ Neu phân tách tuyến tính, các vector hỗ trợ nằm

Select one:

- a. trên đường phân tách âm
- O b. trên đường phân tách dương
- o c. trên đường phân tách âm hoặc đường phân tách dương
- O d. nằm ở vùng giữa đường phân tách âm loặc đường phân tách dương

Clear my choice

Previous page



Previous page

Time left 0:14:16

Question 34

Not yet

Marked out of 1.00 Giả sử bạn huấn luyện một mô hình phân loại trên một tập huấn luyện gồm 10,000 điểm và thu được độ chính xác trên tập đó là 99%. Tuy nhiên, khi bạn gửi nó lên Kaggle thì thu được độ chính xác là 67%. Cách làm nào sau đây có khả năng giúp bạn tăng hiệu quả trên Kaggle?

Select one or more:

- a. Đặt hệ số hiệu chỉnh (nếu có) bằng 0
- ✓ b. Dùng bước tối ưu tham số
- c. Huấn luyện trên nhiều dữ liệu hơn
- d. Bò bớt dữ liệu một cách ngẫu nhiên khi huấn luyện

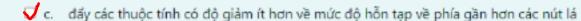
Previous page

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 nút gốc
- □ b. đẩy các thuộc tính có độ giảm nhiều hơn về mức độ hỗn tạp về phía gần hơn các nút lá





Previous page

Not yet answered

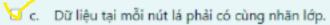
Marked out of 1.00 Những phát biểu nào sau đây là đúng về cây quyết định?

Select one or more:

a. Độ sâu của cây không bao giờ vượt quá O(log n) đối với n ví dụ học.



√ b. Cắt tỉa cây (prunning) thường đạt được độ chính xác tốt hơn là ngừng sớm (early stopping).



d. Chỉ có thể dùng cho bài toán phân loại.

Not yet answered

Marked out of 1.00 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 có thể được dùng để đánh giá hiệu quả và lựa chọn tham số cho một mô hình học máy
- c. Đánh giá chéo không tốn kém 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 có thể tốn thời gian



Not yet answered

Marked out of 1.00

Select one:

K-means là

O a. một thuật toán giám sát làm việc trên dữ liệu có nhãn

O b. một thuật toán giám sát làm việc trên dữ liệu không có nhãn

c. một thuật toán không giám sát làm việc trên dữ liệu không có nhãn

O d. một thuật toán không giám sát làm việc trên dữ liệu có nhãn

Previous page

Not yet answered

Marked 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 (Class label).

	x_1	x_2	Class
2	-1	1	c_2
1	0	1	c_1
ک ح	0	2	c_2
2	1	-1	c_2
1	1	0	c_1
1	1	2	c_1
2	2	2	c_2
3	2	3	c_1

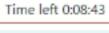
Áp dụng giải thuật học láng giềng gần nhất và sử dụng 7 láng giềng gần nhất, hãy xác định nhãn lớp cho ví dụ ($x_1=1, x_2=1$)?

Select one:



O b. Không xác định

 \odot c. c_1



Not yet answered

Marked out of 1.00 Thuật toán Perceptron sẽ hội tụ

Select one:

✓a. nếu dữ liệu phân tách tuyến tính

- O b. dù cho dữ liệu không phân tách tuyến tính
- O c. nếu chúng ta khởi tạo tất cả các tham số mạng bằng 0
- O d. luôn luôn hội tụ

Previous page

Finish attempt ...