

## Question 1

Not yet  
answeredMarked 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 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

[Next page](#)

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

## Quiz navigation

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

[Finish attempt ...](#)

Time left 0:43:54

## Question 2

Not yet  
answeredMarked out of  
1.00

Trong 2 giai đoạn huấn luyện và dự đoán, thì giải thuật học láng giềng gần nhất tốn chi phí tính toán nhiều hơn cho giai đoạn nào?

Select one:

- ☐ a. Cho giai đoạn dự đoán
- ☒ b. Cho giai đoạn huấn luyện

[Clear my choice](#)[Previous page](#)[Next page](#)

## Question 3

Not yet  
answeredMarked 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 là một
- ☐ b. Ta không thể nói chúng giao nhau hay không giao nhau
- ☒ c. Chúng không giao nhau
- ☐ d. Chúng có thể giao nhau

Handwritten notes in red:

$$f(x) = x$$

$$f(b) = bx$$

$$f(x) = \cancel{bx}$$

$$f(x) = ax + b$$

$$W = W_0$$

$$nx$$

Handwritten notes in blue:

$\mathbf{F}$  : Không gian hàm tuyến tính

$\mathbf{W}$  : Không gian vector tham số

Previous page

Next page

## Question 4

Not yet  
answeredMarked out of  
1.00

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

Select one:

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

[Previous page](#)[Next page](#)

## Question 5

Not yet  
answeredMarked 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 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
- ☐ 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 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](#)[Next page](#)

## Question 6

Not yet  
answeredMarked out of  
1.00

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

Select one or more:

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

[Previous page](#)[Next page](#)

## Question 7

Not yet  
answeredMarked out of  
1.00

Những phát biểu nào sau đây đúng về SVM?

Select one or more:

- ☒ a. Tăng tham số  $C$  hướng tới giảm nhạy cảm với (outliers)
- ☒ b. Tăng tham số  $C$  hướng tới giảm lỗi huấn luyện
- ☒ c. Hard-margin SVM là một trường hợp đặc biệt của soft-margin SVM khi  $C$  được cho bằng 0
- ☒ d. Tăng tham số  $C$  hướng tới giảm lề

[Previous page](#)[Next page](#)

Time left 0:37:09

Question 8

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
- ☐ b. Random forests
- ☐ c. Neural nets
- ☐ d. linear SVM

Previous page

Next page



Time left 0:36:09

Question 9

Not yet  
answered

Marked out of  
1.00

Với một cây quyết định học được, thì nhãn 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 nhỏ nhất trong cây
- ☐ c. đường đi có độ sâu lớn nhất trong cây
- ☐ d. tất cả các đường đi trong cây

Previous page

Next page

Time left 0:36:00

Question **10**

Not yet  
answered

Marked out of  
1.00

Đâu không phải là phương pháp làm sạch dữ liệu?

Select one:

- ☐ a. Phân cụm (Clustering)
- ☐ b. Hồi quy (Regression)
- ☐ c. Phân khoảng (Binning)
- ☒ d. Tích hợp (Aggregation)

Previous page

Next page

Time left 0:35:34

Question 11

Not yet  
answered

Marked out of  
1.00

Mô hình hồi quy sau đây thuộc lớp 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

Previous page

Next page

Time left 0:34:40

Question **12**

Not yet  
answered

Marked out of  
1.00

Khai phá tri thức là ...

Select one:

- ☐ 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
- ☐ b. quá trình chuyển đổi dữ liệu từ dạng này sang một dạng cấu trúc khác
- ☐ 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

Next page

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
- ☐ 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](#)

[Next page](#)

Time left 0:33:57

## Question 14

Not yet  
answeredMarked 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:

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

[Clear my choice](#)

[Previous page](#)[Next page](#)

Time left 0:32:59

Question 15

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. Lý do chính là thời gian tính toán lớn
- ☐ 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.

Previous page

Next page

## Question 16

Not yet  
answeredMarked 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.

[Previous page](#)[Next page](#)



## Question 17

Not yet  
answeredMarked out of  
1.00

Kỹ thuật nào dưới đây có thể giúp giảm Overfitting?

Select one:

- ☐ a. Bỏ bớt dữ liệu nếu nó có kích cỡ quá lớn
- ☐ b. Thay đổi mô hình mới hoặc phương pháp học mới
- ☒ c. Sử dụng Hiệu chỉnh (regularization), nhằm giới hạn lại không gian tìm kiếm của hàm cần học

[Previous page](#)[Next page](#)

Time left 0:29:52

Question **18**

Not yet  
answered

Marked out of  
1.00

Xác suất hậu nghiệm (Posterior probability) nói đến

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
- ☐ c. tri thức của một mô hình

[Clear my choice](#)

[Previous page](#)

[Next page](#)

Time left 0:29:45

Question **19**

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

Previous page

Next page

## Question 20

Not yet  
answeredMarked 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:

- ☒ a. Recall cho từng lớp
- ☒ b. Precision cho từng lớp
- ☐ c. Trung bình bình phương lỗi (Mean squared error)
- ☐ d. Accuracy

[Previous page](#)[Next page](#)

Time left 0:27:21

Question **21**

Not yet  
answered

Marked out of  
1.00

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
- ☒ b. Tối ưu một hàm mục tiêu tuyến tính
- ☒ 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

Next page

## Question 22

Not yet  
answeredMarked 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 suất đầy đủ khi muốn đưa ra phán đoán nhãn 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](#)[Next page](#)

Time left 0:25:35

Question **23**

Not yet  
answered

Marked out of  
1.00

Phát biểu nào sau đây là SAI?

Select one:

- ☐ a. Lựa chọn mô hình là một bước bắt buộc khi muốn so sánh nhiều mô hình (hoặc phương pháp) học máy khác nhau
- ☒ b. Đánh giá mô hình và lựa chọn mô hình trong Học máy là hai thứ độc lập với nhau
- ☐ c. Đánh giá mô hình thường yêu cầu thực hiện bước lựa chọn mô hình

[Previous page](#)

[Next page](#)

## Question 24

Not yet  
answeredMarked out of  
1.00

Tiền xử lý dữ liệu có thể xem như

Select one:

- ☒ a. là quá trình biến đổi dữ liệu thô tới một dạng thích hợp để cho bước phân tích
- ☐ b. phân tích một số tính chất chính của tập dữ liệu
- ☐ c. là quá trình quyết định định dạng về dữ liệu và thu thập dữ liệu

[Previous page](#)[Next page](#)



Time left 0:24:21

Question **25**

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. cắt tia
- ☐ d. tìm các tập mục thường xuyên

Previous page

Next page

## Question 26

Not yet  
answeredMarked out of  
1.00

Những công việc nào dưới đây mà sử dụng phân cụm có thể là cách tiếp cận phù hợp?

Select one or more:

- ☐ a. Cho trước dữ liệu lịch sử thời tiết, hãy dự đoán xem thời tiết ngày mai là nắng hay mưa.
- ☐ b. Cho trước dữ liệu bán hàng từ một số lượng lớn sản phẩm trong siêu thị, hãy dự đoán doanh số bán hàng trong tương lai cho từng sản phẩm này.
- ☒ c. Cho trước cơ sở dữ liệu về người dùng, tự động nhóm họ thành các phân khúc thị trường khác nhau.
- ☒ d. Từ các kiểu sử dụng của người dùng trên một trang web, hãy xác định các nhóm người dùng khác nhau.

[Previous page](#)[Next page](#)

## Question 27

Not yet  
answeredMarked 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:

- ☒ a.  $w_1$  có thể không bằng  $w_2$ .
- ☐ b.  $S_1$  có thể không là con của  $S_2$
- ☒ c.  $S_1 \subset S_2$
- ☐ d.  $w_1 = w_2$

[Previous page](#)[Next page](#)

## Question 28

Not yet  
answeredMarked 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 thử
- ☒ 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](#)[Next page](#)

## Question 29

Not yet  
answeredMarked 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 huấn 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 (underfitting) 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](#)[Next page](#)

Time left 0:18:04

Question **30**

Not yet  
answered

Marked out of  
1.00

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 và cùng số lượng đầu vào?

Select one:

☒ True

☐ False

Previous page

Next page

Time left 0:16:51

Question 31

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

Previous page

Next page

## Question 32

Not yet  
answeredMarked out of  
1.00

Trong hard-margin SVM cho dữ liệu phân tách tuyến tính, các vector hỗ trợ nằm

Select one:

- ☐ a. trên đường phân tách âm
- ☐ b. trên đường phân tách dương
- ☒ c. trên đường phân tách âm hoặc đường phân tách dương
- ☐ d. nằm ở vùng giữa đường phân tách âm hoặc đường phân tách dương

[Clear my choice](#)

[Previous page](#)[Next page](#)



Time left 0:15:04

Question **33**

Not yet  
answered

Marked out of  
1.00

Thuật toán K-means cho phân cụm đảm bảo hội tụ tới một điểm cực trị địa phương hoặc điểm dừng?

Select one:

☒ True

☐ False

Previous page

Next page

## Question 34

Not yet  
answeredMarked 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. Bỏ bớt dữ liệu một cách ngẫu nhiên khi huấn luyện

[Previous page](#)[Next page](#)

Time left 0:13:54

Question 35

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

Previous page

Next page

## Question 36

Not yet  
answeredMarked 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 (pruning) thường đạt được độ chính xác tốt hơn là ngừng sớm (early stopping).
- ☒ c. Dữ liệu tại mỗi nút lá phải có cùng nhãn lớp.
- ☐ d. Chỉ có thể dùng cho bài toán phân loại.

[Previous page](#)[Next page](#)

## Question 37

Not yet  
answeredMarked 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

Time left 0:10:22

Question **38**

Not yet  
answered

Marked out of  
1.00

K-means là

Select one:

- ☐ a. một thuật toán giám sát làm việc trên dữ liệu có nhãn
- ☐ 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
- ☐ d. một thuật toán không giám sát làm việc trên dữ liệu có nhãn

Previous page

Next page

## Question 39

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 (Class label).

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

Select one:

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

Time left 0:08:43

Question **40**

Not yet  
answered

Marked out of  
1.00

Thuật toán Perceptron sẽ hội tụ

Select one:

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

Previous page

Finish attempt ...