

Câu 5. Khi phương sai (*variance*) của dữ liệu là 0, dữ liệu có đặc điểm gì?

- a. Không có phân tán trong dữ liệu.
- b. Các phần tử dữ liệu như nhau.
- c. Không có ngoại biên trong dữ liệu.
- d. Câu a, b, và c đều đúng.
- e. Ý kiến khác.

Bảng 1. Dữ liệu trong không gian 2 chiều (X_1, X_2)

OID	X_1	X_2
1	2.5	2
2	3	4
3	7	6.8
4	8.3	9
5	5.4	5.2
6	6.7	10
7	9	9
8	4	4.1
9	7.8	8
10	10	8.5

Câu 6. Xác định vector trung bình của các đối tượng dữ liệu trong không gian 2 chiều ở **Bảng 1**. Lưu ý các kết quả làm tròn với 2 số thập phân.

- a. (5.40, 5.20)
- b. (6.37, 6.66)
- c. (6.70, 10.00)
- d. (6.85, 7.40)
- e. Ý kiến khác.

Câu 7. Cho biến X_2 ở **Bảng 1** có $Q1 = 4.05$ và $Q2 = 8.75$, xác định các phần tử có giá trị nhiều ở biến X_2 .

- a. Phần tử $OID = 1$
- b. Phần tử $OID = 10$
- c. Phần tử $OID = 1$ và phần tử $OID = 10$
- d. Không có phần tử có giá trị nhiều ở biến X_2 .
- e. Ý kiến khác.

Bảng 2. Mẫu trong không gian 2 chiều (X_1, X_2)

OID	X_1	X_2
2	3	4
8	4	4.1
9	7.8	8

Câu 8. Thực hiện lấy mẫu 30% từ tập các đối tượng dữ liệu trong **Bảng 1** được kết quả mẫu ở **Bảng 2**. Cho biết cơ chế lấy mẫu đã thực hiện là cơ chế nào?

- a. Lấy mẫu không đặt lại (Simple random sampling without replacement)
- b. Lấy mẫu có đặt lại (Simple random sampling with replacement)
- c. Đều có thể thực hiện cơ chế lấy mẫu không đặt lại hoặc lấy mẫu có đặt lại để có **Bảng 2**.

- d. Câu a, b, và c đều sai do không đủ chi tiết để xác định cơ chế lấy mẫu.
e. Ý kiến khác.

Câu 9. Khi thực hiện dự báo với kỹ thuật hồi quy tuyến tính đa biến, giả định nào được yêu cầu trên dữ liệu huấn luyện?

- a. Các biến độc lập không có tương quan với nhau.
b. Lỗi dự báo cho mỗi đối tượng tuân theo phân phối chuẩn, có phương sai không đổi.
c. Lỗi dự báo cho mỗi đối tượng không có tương quan với nhau.
d. Câu a, b, và c đều đúng.
e. Ý kiến khác.

Câu 10. Kết quả nào sau đây là kết quả của kỹ thuật hồi quy tuyến tính đa biến dùng cho việc dự báo giá trị của biến y?

- a. $y = x_1^2 - 3x_2 + 0.5$**
b. $y = x_1x_2 + 1.5$
c. $y = x_1 + 1.5/x_2 + x_3^{x_1x_2} - 2$
d. Câu a, b, và c đều sai.
e. Ý kiến khác.

Câu 11. Giả sử có 2 kết quả cho hồi quy tuyến tính đơn biến để dự báo giá trị của biến X_2 từ giá trị của biến X_1 trong **Bảng 2** như sau:

Kết quả 1: $X_2 = 0.93X_1 + 0.75$

Kết quả 2 : $X_2 = X_1 + 0.5$

Hãy cho biết dự báo với kết quả 1 tốt hơn hay dự báo với kết quả 2 tốt hơn ?

- a. Kết quả 1**
b. Kết quả 2
c. Kết quả dự báo như nhau
d. Không thể xác định do thiếu chi tiết
e. Ý kiến khác.

Câu 12. Cho kết quả hồi quy tuyến tính đa biến để dự báo giá trị của biến y như sau:

$$Y = X_1 - 2.5X_2 + 0.7X_3 - 4.1$$

Chọn phát biểu ĐÚNG về sự liên hệ giữa các biến dự báo và biến đáp ứng?

- a. X_1 có ảnh hưởng đến Y nhiều nhất.
b. X_2 tương quan nghịch với Y.
c. X_3 đóng góp vào kết quả dự báo của Y bằng với X_1 .
d. Câu a, b, và c đều đúng.
e. Ý kiến khác.

Câu 13. Chọn phát biểu ĐÚNG về bài toán phân lớp so với bài toán gom cụm và hồi quy.

- a. Phân lớp là bài toán dự báo các nhãn mô tả lớp của các đối tượng dữ liệu dựa trên các nhãn mô tả lớp được xác định trước của các đối tượng dữ liệu khác.**

- b. Phân lớp là bài toán nhóm các đối tượng dữ liệu vào các nhóm tương tự nhau chưa được xác định trước.
- c. Phân lớp là bài toán dự báo giá trị số của các thuộc tính của các đối tượng dữ liệu dựa trên các giá trị sẵn có của các thuộc tính của các đối tượng dữ liệu khác.
- d. Câu a, b, và c đều sai.
- e. Ý kiến khác.

Câu 14. Kết quả phân lớp nào sau đây là phân lớp phi tuyến?

- a. Cây quyết định C4.5
- b. Naïve Bayes
- c. Mạng nơron nhân tạo với hàm hoạt tính *sigmoid* cho các nút
- d. Câu a, b, và c đều sai.
- e. Ý kiến khác.

Câu 15. Nếu k -nn được dùng để phân lớp các đối tượng trong không gian của các đối tượng trong **Bảng 1** thì công tác gì cần được thực hiện?

- a. Độ đo sai biệt giữa các đối tượng trong không gian này cần được định nghĩa.
- b. Chỉ mục các đối tượng trong bảng dữ liệu cần được xây dựng.
- c. Các phần tử nhiễu cần được xác định và xử lý phù hợp.
- d. Câu a, b, và c đều đúng.
- e. Ý kiến khác.

Câu 16. Cho phương trình của mặt phẳng phân tách đạt được trong không gian biến đổi với *support vector machines* như sau: $-6.64 \cdot x_1 - 9.32 \cdot x_2 + 7.93 = 0$. Xác định lớp của đối tượng $X = (1.20, 0.43)$.

- a. Đối tượng X là đối tượng nhiễu nên không thuộc về lớp 1, cũng không thuộc về lớp -1.
- b. Lớp 1.
- c. Lớp -1.
- d. Không đủ thông tin để xác định đối tượng X thuộc về lớp 1 hay lớp -1.
- e. Ý kiến khác.

Câu 17. Giả sử mỗi đối tượng ở **Bảng 1** có thêm thuộc tính lớp Y với 2 giá trị lớp là 0 và 1. Nếu thực hiện phân lớp với mạng nơron nhân tạo thì mạng nơron này có tầng nhập và tầng xuất là gì?

- a. Tầng nhập có 3 nút tương ứng cho thuộc tính OID, X_1 , X_2 và tầng xuất có 2 nút tương ứng cho lớp 0 và lớp 1.
- b. Tầng nhập có 3 nút tương ứng cho thuộc tính OID, X_1 , X_2 và tầng xuất có 1 nút tương ứng cho thuộc tính lớp Y .
- c. Tầng nhập có 2 nút tương ứng cho thuộc tính X_1 , X_2 và tầng xuất có 2 nút tương ứng cho thuộc tính OID và thuộc tính lớp Y .
- d. Tầng nhập có 2 nút tương ứng cho thuộc tính X_1 , X_2 và tầng xuất có 1 nút tương ứng cho thuộc tính lớp Y .
- e. Ý kiến khác.

Bảng 3. Dữ liệu về khách hàng

OID	Transaction	Value	Class
1	FREQUENT	HIGH	A
2	NORMAL	MEDIUM	B
3	FREQUENT	LOW	B
4	NORMAL	HIGH	A
5	INFREQUENT	MEDIUM	B
6	INFREQUENT	LOW	B
7	NORMAL	HIGH	A
8	FREQUENT	MEDIUM	A

Câu 18. Cho dữ liệu về khách hàng trong **Bảng 3**. Nếu xây dựng cây quyết định CART để thực hiện phân lớp khách hàng thành A và B thì thuộc tính nào sẽ được chọn làm nút gốc?

- a. Transaction với {FREQUENT} và {NORMAL, INFREQUENT}
- b. Transaction với {INFREQUENT} và {FREQUENT, NORMAL}
- c. Value với {HIGH} và {MEDIUM, LOW}
- d. Value với {LOW} và {HIGH, MEDIUM}
- e. Ý kiến khác.

Câu 19. Cho dữ liệu về khách hàng trong **Bảng 3**. Giả sử các thuộc tính độc lập với nhau cho từng lớp khách hàng A hoặc B. Thực hiện phân lớp với Naïve Bayes thì lớp của khách hàng (INFREQUENT, HIGH) là gì?

- a. Lớp A
- b. Lớp B
- c. Chưa đủ thông tin để xác định lớp của khách hàng này.
- d. Câu a, b, và c đều sai.
- e. Ý kiến khác.

Câu 20. Cho các ma trận nhầm lẫn từ 2 giải thuật phân lớp M1 và M2 như sau. Xác định kết quả phân lớp hiệu quả hơn dựa vào độ chính xác (*accuracy*).

Ma trận nhầm lẫn của bộ phân lớp M1

	TN	CTN	CC1	CC2	BTH
TN	268	0	0	0	0
CTN	15	246	4	1	0
CC1	0	0	268	0	0
CC2	0	0	2	258	8
BTH	0	13	4	6	245

Ma trận nhầm lẫn của bộ phân lớp M2

	TN	CTN	CC1	CC2	BTH
TN	268	0	0	0	0
CTN	16	235	1	0	14
CC1	0	0	268	0	0
CC2	0	0	4	264	0
BTH	0	27	11	6	224

- a. M1 và M2 hiệu quả như nhau.
- b. M1 hiệu quả hơn M2.
- c. M2 hiệu quả hơn M1.
- d. Không đủ thông tin để so sánh hiệu quả của M1 và M2.
- e. Ý kiến khác.

Câu 21. Trong số các giải thuật gom cụm sau, giải thuật nào chịu ảnh hưởng của nhiễu nhiều nhất?

- a. *k*-means
- b. PAM
- c. HAC
- d. DBSCAN
- e. Ý kiến khác.

Câu 22. Trong số các giải thuật gom cụm sau, giải thuật nào có thể khám phá ra các cụm có hình dạng tùy ý, tùy vào đặc điểm của dữ liệu?

- a. *k*-means
- b. PAM
- c. HAC
- d. DBSCAN
- e. Ý kiến khác.

Câu 23. Chọn phát biểu ĐÚNG về tác vụ gom cụm dữ liệu.

- a. Gom cụm giúp hình thành nên các nhóm đối tượng tương tự nhau.
- b. Gom cụm là trường hợp đặc biệt của phân lớp khi tất cả các đối tượng đều chưa có nhãn mô tả về nhóm của mỗi đối tượng.
- c. Gom cụm đưa các đối tượng giống nhau về các cụm kết quả và đưa các đối tượng khác nhau nhất về các cụm nhiễu.
- d. Câu a, b, và c đều đúng.
- e. Ý kiến khác.

Câu 24. Thực hiện gom cụm dữ liệu trên các đối tượng ở **Bảng 1** bằng giải thuật *k*-means với $k=2$ và độ đo Euclidean. Sau 1 số lần lặp, trung tâm cụm của cụm 1 là: $c_1 = (3.73, 3.83)$ và trung tâm cụm của cụm 2 là: $c_2 = (8.13, 8.55)$. Xác định liệu đối tượng $OID = 3$ gần trung tâm cụm hơn hay là gần biên hơn so với đối tượng $OID = 6$?

- a. Đối tượng $OID = 3$ gần trung tâm cụm hơn đối tượng $OID = 6$.
- b. Đối tượng $OID = 3$ gần biên hơn so với đối tượng $OID = 6$.
- c. Đối tượng $OID = 3$ gần trung tâm cụm như đối tượng $OID = 6$.
- d. Câu a và b đều sai do không có chi tiết để xác định sự liên hệ giữa hai đối tượng này.
- e. Ý kiến khác.

Câu 25. Thực hiện gom cụm dữ liệu trên các đối tượng ở **Bảng 1** bằng giải thuật DBSCAN với $\epsilon = 2.24$ và $MinPts = 2$. Danh sách các đối tượng lõi cùng với láng giềng như sau :

Đối tượng $OID = 2$ có láng giềng : $OID = 1, OID = 8$

Đối tượng $OID = 4$ có láng giềng : $OID = 6, OID = 7, OID = 9, OID = 10$

Đối tượng $OID = 7$ có láng giềng : $OID = 4, OID = 9, OID = 10$

Đối tượng $OID = 8$ có láng giềng : $OID = 2, OID = 5$

Đối tượng $OID = 9$ có láng giềng : $OID = 3, OID = 4, OID = 7$

Đối tượng $OID = 10$ có láng giềng : $OID = 4, OID = 7$

Xác định đặc điểm cụm của đối tượng $OID = 5$ và đối tượng $OID = 6$.

- a. Đối tượng OID = 5 và đối tượng OID = 6 thuộc về cùng cụm của đối tượng lỗi OID = 4, cũng là cụm của đối tượng lỗi OID = 8.
- b. Đối tượng OID = 5 thuộc về cụm của đối tượng lỗi OID = 8; trong khi đó, không thuộc về cùng cụm với đối tượng OID = 5, đối tượng OID = 6 thuộc về cụm của đối tượng lỗi OID = 4.
- c. Đối tượng OID = 5 và đối tượng OID = 6 đều là đối tượng nhiễu.
- d. Câu a, b, và c đều sai do thiếu chi tiết xác định cụm cho các đối tượng này.
- e. Ý kiến khác.

Câu 26. Cho luật kết hợp như sau : {mua_máy_tính, mua_windows} \rightarrow {mua_antivirus} với support = 0.8 và confidence = 1. Chọn phát biểu ĐÚNG về luật kết hợp này.

- a. Tất cả khách hàng đều mua cùng lúc máy tính, windows, và antivirus.
- b. Tất cả khách hàng đều mua cùng lúc máy tính và windows.
- c. Tất cả khách hàng mua cùng lúc máy tính và windows cũng đều mua antivirus.
- d. Câu a, b, và c đều sai.
- e. Ý kiến khác.

Câu 27. Giải thuật Apriori sử dụng ràng buộc nào sau đây để tiết kiệm chi phí ?

- a. Nếu tập phần tử cha không thỏa ngưỡng hỗ trợ thì tất cả các tập phần tử con đều không thỏa ngưỡng hỗ trợ.
- b. Nếu tập phần tử cha ở mức $(k+1)$ không thỏa ngưỡng hỗ trợ thì chỉ tất cả các tập phần tử con ở mức (k) đều không thỏa ngưỡng hỗ trợ.
- c. Nếu tập phần tử con không thỏa ngưỡng hỗ trợ thì tất cả các tập phần tử cha đều không thỏa ngưỡng hỗ trợ.
- d. Nếu tập phần tử con ở mức (k) không thỏa ngưỡng hỗ trợ thì chỉ tất cả các tập phần tử cha ở mức $(k+1)$ đều không thỏa ngưỡng hỗ trợ.
- e. Ý kiến khác.

Câu 28. Cho tập các giao tác trong cơ sở dữ liệu về các điểm du lịch mà khách hàng đã đến thăm trong **Bảng 4** như sau :

Bảng 4. Cơ sở dữ liệu về các điểm du lịch

TID	Địa điểm
1	A1, A2
2	A2, A3, A5, A6
3	A1, A4
4	A1, A3, A4, A5
5	A2, A5, A6
6	A1, A3, A4
7	A2, A3, A4, A5
8	A1, A2, A6

Thực hiện khai phá các tập mẫu phổ biến với ngưỡng đếm hỗ trợ là 2 (min_sup_count = 2) dùng giải thuật Apriori. Giả sử giải thuật đang thực thi đến bước tạo các ứng cử viên từ tập L2 như sau :

$L2 = \{\{A1, A2\} \text{ với support} = 2, \{A1, A3\} \text{ với support} = 2, \{A1, A4\} \text{ với support} = 3, \{A2, A3\} \text{ với support} = 2, \{A2, A5\} \text{ với support} = 3, \{A2, A6\} \text{ với support} = 3, \{A3, A4\} \text{ với support} = 3, \{A3, A5\} \text{ với support} = 3, \{A4, A5\} \text{ với support} = 2, \{A5, A6\} \text{ với support} = 2\}$

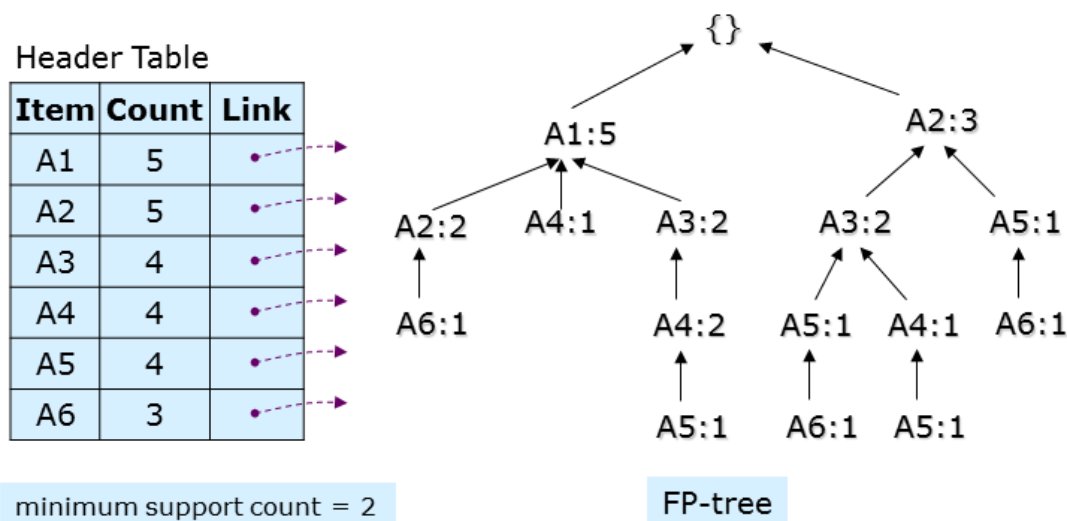
Xác định tập mẫu phổ biến $L3$ ở mức kế tiếp ?

- $L3 = \{\{A1, A3, A4\} \text{ với support} = 2, \{A2, A3, A5\} \text{ với support} = 2, \{A2, A5, A6\} \text{ với support} = 2, \{A3, A4, A5\} \text{ với support} = 2\}$
- $L3 = \{\{A1, A3, A4\} \text{ với support} = 2, \{A2, A3, A5\} \text{ với support} = 2, \{A2, A4, A5\} \text{ với support} = 2, \{A2, A5, A6\} \text{ với support} = 2, \{A3, A4, A5\} \text{ với support} = 2\}$
- $L3 = \{\{A1, A2, A6\} \text{ với support} = 2, \{A1, A3, A4\} \text{ với support} = 2, \{A2, A3, A5\} \text{ với support} = 2, \{A2, A5, A6\} \text{ với support} = 2, \{A3, A4, A5\} \text{ với support} = 2\}$
- $L3 = \{\{A1, A3, A4\} \text{ với support} = 2, \{A2, A5, A6\} \text{ với support} = 2, \{A3, A4, A5\} \text{ với support} = 2\}$
- Ý kiến khác.

Câu 29. Cây mẫu phổ biến FP-tree có đặc điểm gì sau đây?

- Cây nhiều nhánh với mỗi nhánh ứng với một tập phân tử phổ biến trong cơ sở dữ liệu.
- Cây nhiều nhánh với mỗi nhánh là kết quả tích lũy từ những giao tác chứa những phân tử phổ biến tương ứng trong cơ sở dữ liệu.
- Cây nhiều nhánh không có nhãn ở nút gốc và các nút lá là các nút được liên kết từ các phân tử trong bảng tiêu đề (*header table*).
- Câu a, b, và c đều sai.
- Ý kiến khác.

Câu 30. Cho cây FP-tree cùng với bảng tiêu đề như sau khi thực hiện khai phá các tập mẫu phổ biến với ngưỡng đếm hỗ trợ là 2 ($\text{min_sup_count} = 2$) dùng giải thuật FP-Growth trên dữ liệu ở **Bảng 4**. Cho đơn giản, chi tiết của liên kết từ bảng tiêu đề đến cây được lược bỏ.



Xác định các tập phân tử phổ biến được dẫn ra từ phân tử $A6$?

- $\{A6\}$ với support = 3, $\{A6, A2\}$ với support = 2, $\{A6, A5\}$ với support = 2, $\{A6, A1, A2\}$ với support = 1, $\{A6, A2, A3, A5\}$ với support = 2, $\{A6, A2, A5\}$ với support = 2

- b. $\{A6\}$ với $\text{support} = 3$, $\{A6, A2\}$ với $\text{support} = 2$, $\{A6, A5\}$ với $\text{support} = 2$, $\{A6, A2, A5\}$ với $\text{support} = 2$
- c. $\{A6\}$ với $\text{support} = 3$, $\{A6, A2\}$ với $\text{support} = 2$, $\{A6, A5\}$ với $\text{support} = 2$, $\{A6, A2, A5\}$ với $\text{support} = 2$, $\{A6, A2, A3, A5\}$ với $\text{support} = 2$
- d. Câu a, b, và c đều sai.
- e. Ý kiến khác.

Câu 31. Tự luận về việc so sánh giải thuật Apriori và giải thuật FP-Growth có giải thích với các tiêu chí so sánh sau:

31.1. Cách tiếp cận (1 điểm)

31.2. Số lần duyệt qua cơ sở dữ liệu (0.5 điểm)

Câu 32. Tự luận về phân lớp dữ liệu với mạng nơron nhân tạo và *support vector machines*.

Cho dữ liệu trong không gian 1 chiều như sau :

$X_1 = (0)$ với lớp 1

$X_2 = (-1)$ với lớp -1

32.1. Giả sử dùng hàm: $\phi(X) = (x^2, \sqrt{2}x, -1)$ để biến đổi các đối tượng dữ liệu trong không gian 1 chiều đã cho vào không gian biến đổi 3 chiều. Xác định hàm kernel $K(X_1, X_2)$ với $X_1 = (x_{1,1})$ và $X_2 = (x_{2,1})$ có thể sử dụng thay cho tích nội $\phi(X_1)^T \phi(X_2)$ với hàm biến đổi tương minh $\phi(X)$ này. (0.5 điểm)

32.2. Giả sử sử dụng *support vector machines* để xây dựng hàm phân lớp cho các đối tượng dữ liệu trong không gian biến đổi 3 chiều. Cho trước nhân tử Lagrange tương ứng X_1 và X_2 trong dữ liệu huấn luyện đã cho là : $\alpha_1 = \alpha_2 = 2/3$, tìm vector pháp tuyến w và độ dời b trong không gian biến đổi 3 chiều, và từ đó, tìm hàm phân lớp cho các đối tượng dữ liệu trong không gian biến đổi 3 chiều. (1.25 điểm)

Điểm thưởng : vẽ các biên phân lớp cho các đối tượng dữ liệu trong không gian gốc 1 chiều. (1 điểm)

32.3. Nếu dùng mạng nơron nhân tạo có dạng là mạng đa tầng truyền thẳng để phân lớp các đối tượng dữ liệu trong không gian biến đổi 3 chiều với hàm biến đổi $\phi(X)$ ở trên thì cấu hình mạng cho giải thuật học *backpropagation* nên là gì? Giải thích? (0.75 điểm)

<u>Khoa/Bộ Môn</u> :	<u>Giảng viên ra đề</u> : TS. Võ Thị Ngọc Châu

Họ - Tên:

Mã Số Sinh Viên:

Mã môn học: **CO3029**

Mã đề: **F17101**

Ngày kiểm tra: **26/12/2017**

Phần trả lời

Câu 1 - 15:

Câu	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
a										X	X		X		
b	X		X			X						X			
c								X						X	
d		X		X	X		X		X						X
e															

Câu 16 - 30:

Câu	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
a				X		X		X					X		
b					X				X	X				X	X
c	X		X								X	X			
d		X					X								
e															

Câu 7. $Q1 = 4.05$ và $Q3 = 8.75$

Câu 30. b/e. $\{A6\}$ với support = 3, $\{A6, A2\}$ với support = 3, $\{A6, A5\}$ với support = 2, $\{A6, A2, A5\}$ với support = 2

Câu 31. Tự luận:

Tiêu chí	Apriori	FP-Growth
31.1. Cách tiếp cận	Giải thuật lập để khám phá các tập phần tử phổ biến từng mức dùng tri thức biết trước về sự liên hệ giữa tính phổ biến của tập phần tử cha và con nhằm hạn chế số ứng cử viên được sinh ra và kiểm tra. Sử dụng danh sách để tổ chức dữ liệu và kết quả.	Giải thuật đệ quy, cơ chế chia-đề-trị, thực hiện tổ chức lại dữ liệu với cây mẫu phổ biến FP-tree và sau đó thực hiện phát triển các tập mẫu phổ biến từ cây FP-tree nên không duyệt đi duyệt lại cơ sở dữ liệu và không dùng cơ chế sinh ứng cử viên và kiểm tra. Sử dụng cấu trúc cây để tổ chức dữ liệu và kết quả.
31.2. Số lần duyệt cơ sở dữ liệu	$(k+1)$ với k là chiều dài tập phần tử phổ biến dài nhất vì khi L_{k+1} rỗng, giải thuật sẽ ngưng thực thi và ứng với mỗi C_{k+1} là 1 lần duyệt cơ sở dữ liệu	2 với 1 lần cho việc xây dựng bảng tiêu đề L_1 và 1 lần cho việc xây dựng cây FP-tree từ việc nén cơ sở dữ liệu

Họ - Tên:

Mã Số Sinh Viên:

Mã môn học: **CO3029**

Mã đề: **F17101**

Ngày kiểm tra: **26/12/2017**

Câu 32. Tự luận:

32.1. Hàm kernel:

$$K(X_1, X_2) = (x_{1,1}^2, \sqrt{2}.x_{1,1}, -1)^T (x_{2,1}^2, \sqrt{2}.x_{2,1}, -1)$$

$$K(X_1, X_2) = x_{1,1}^2 x_{2,1}^2 + 2 x_{1,1} x_{2,1} + 1 = (x_{1,1} x_{2,1} + 1)^2$$

32.2. Cho trước nhân tử Lagrange $\alpha_1 = \alpha_2 = 2/3$, tính w , b , $f(X)$.

$$w = \alpha_1 c_1 \phi(X_1)^T + \alpha_2 c_2 \phi(X_2)^T = (2/3)(0, 0, -1)^T - (2/3)(1, -\sqrt{2}, -1)^T$$

$$w = (-2/3, 2\sqrt{2}/3, 0)^T$$

$$b = 1 - w^T \phi(X_1) = 1 - (-2/3, 2\sqrt{2}/3, 0)^T (0, 0, -1) = 1$$

$$f(X) = w^T \phi(X) + b = (-2/3, 2\sqrt{2}/3, 0)^T (x^2, \sqrt{2}x, -1) + 1$$

$$f(X) = (-2/3)x^2 + (4/3)x + 1$$

Nếu $f(X) \geq 1$ thì X thuộc lớp 1.

$$\text{Tương đương: } (-2/3)x^2 + (4/3)x \geq 0$$

$$\text{Tương đương: } x(2-x) \geq 0$$

$$\text{Tương đương: } 0 \leq x \leq 2$$

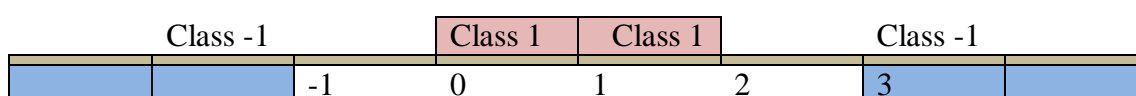
Nếu $f(X) \leq -1$ thì X thuộc lớp -1.

$$\text{Tương đương: } x^2 - 2x - 3 \geq 0$$

$$\text{Tương đương: } (x+1)(x-3) \geq 0$$

$$\text{Tương đương: } x \leq -1 \vee x \geq 3$$

Biên phân lớp trong không gian gốc 1 chiều:



32.3. Cấu hình mạng nơron đa tầng truyền thẳng:

Tầng nhập từ không gian gốc 1 chiều: 1 nút.

Tầng ẩn từ không gian gốc 1 chiều đến không gian biến đổi 3 chiều: 3 nút với các trọng số kết nối là hệ số ở chiều tương ứng: 1, $\sqrt{2}$, -1. Hàm hoạt tính là hàm biến đổi tại chiều tương ứng: x^2 , x , 1.

Tầng ẩn cho giải thuật học: số lượng tầng và số nút tùy vào độ phức tạp của hàm sẽ được học, ví dụ: 1 tầng với số nút là 2, dùng hàm hoạt tính là sigmoid.

Tầng xuất cho phân lớp nhị phân: 1 nút, dùng hàm hoạt tính là sigmoid.