



**TRƯỜNG ĐẠI HỌC TÀI NGUYÊN VÀ MÔI TRƯỜNG
THÀNH PHỐ HỒ CHÍ MINH**

PHIẾU NHẬP CÂU HỎI TRẮC NGHIỆM KHÁCH QUAN

Đơn vị: Khoa Hệ thống Thông tin và Viễn thám

Học phần: Khai phá dữ liệu

Số đơn vị học phần: 02

Trình độ đào tạo: Đại học

Thời gian nhập:

Câu hỏi số: 01

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Hiểu

Mức độ: Dễ

Phần nội dung câu hỏi:

Mục đích của khai phá dữ liệu là gì?

Các đáp án:

| | |
|---|-----------------------------------------------------------------------------------------------|
| A | Sự gia tăng bùng nổ của dữ liệu: Từ mức độ Terabytes đến mức độ Petabytes. |
| B | Chúng ta bị tràn ngập trong dữ liệu – Nhưng lại thiếu (cần) tri thức. |
| C | Khai phá dữ liệu giúp tự động phân tích các tập dữ liệu rất lớn, để khám phá ra các tri thức. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 02

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

Dữ liệu thường được thu thập bằng những công cụ nào?

Các đáp án:

| | |
|---|---------------------------------------|
| A | Các công cụ thu thập dữ liệu tự động. |
| B | Các hệ thống cơ sở dữ liệu. |
| C | World Wide Web, xã hội số. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 03

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

Dữ liệu thường được thu thập từ những nguồn nào?

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------|
| A | Kinh doanh: Internet, thương mại điện tử, giao dịch thương mại, chứng khoán,... |
| B | Khoa học: Tín hiệu cảm biến, tin sinh, thí nghiệm mô phỏng/giả lập,... |

| | |
|---|-----------------------------------------------|
| C | Xã hội: Tin tức, máy ảnh số, các mạng xã hội. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 04

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Hiểu

Mức độ: Trung bình

Phân nội dung câu hỏi:

| |
|-------------------------------------|
| Định nghĩa: Khai phá dữ liệu là gì? |
|-------------------------------------|

Các đáp án:

| | |
|---|------------------------------------------------------------------------------------------------|
| A | Là việc tìm kiếm thông tin (Information retrieval). |
| B | Là việc trích rút ra được các mẫu hoặc tri thức quan trọng từ một lượng lớn dữ liệu (rất) lớn. |
| C | Là việc xử lý các câu truy vấn (SQL) đối với các cơ sở dữ liệu. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: B

Câu hỏi số: 05

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Biết

Mức độ: Khó

Phân nội dung câu hỏi:

| |
|---------------------------------------------|
| Khai phá dữ liệu còn có tên gọi khác là gì? |
|---------------------------------------------|

Các đáp án:

| | |
|---|-----------------------------------------------------------------------|
| A | Khám phá tri thức trong CSDL (Knowledge discovery in database - KDD). |
| B | Trích rút tri thức (Knowledge extraction) |
| C | Phân tích mẫu/dữ liệu (Data/pattern analysis) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 06

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Phân tích

Mức độ: Khó

Phân nội dung câu hỏi:

| |
|-------------------------------------------------------------------------------------------------------------|
| Theo quan điểm về các hệ thống dữ liệu và kho dữ liệu. Quá trình khám phá tri thức nào sau đây là tốt nhất? |
|-------------------------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------------------------------|
| A | Data warehouse -> Data cleaning (Databases) -> Task-relevant Data -> Data Mining -> Pattern Evaluation. |
| B | Data cleaning (Databases) -> Data warehouse -> Data Mining -> Task-relevant Data -> Pattern Evaluation. |
| C | Data cleaning (Databases) -> Task-relevant Data -> Data warehouse -> Data Mining -> Pattern Evaluation. |
| D | Data cleaning (Databases) -> Data warehouse -> Task-relevant Data -> Data Mining -> Pattern Evaluation. |

Đáp án đúng: D

Câu hỏi số: 07

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Phân tích

Mức độ: Khó

Phản nội dung câu hỏi:

Theo quan điểm về học máy và thống kê. Quá trình khám phá tri thức nào sau đây là đúng nhất?

Các đáp án:

| | |
|---|---------------------------------------------------------------------|
| A | Input Data -> Data Pre-Processing -> Data Mining -> Post-Processing |
| B | Data Pre-Processing -> Input Data -> Post-Processing -> Data Mining |
| C | Data Pre-Processing -> Data Mining -> Input Data -> Post-Processing |
| D | Input Data -> Data Pre-Processing -> Post-Processing -> Data Mining |

Đáp án đúng: A

Câu hỏi số: 08

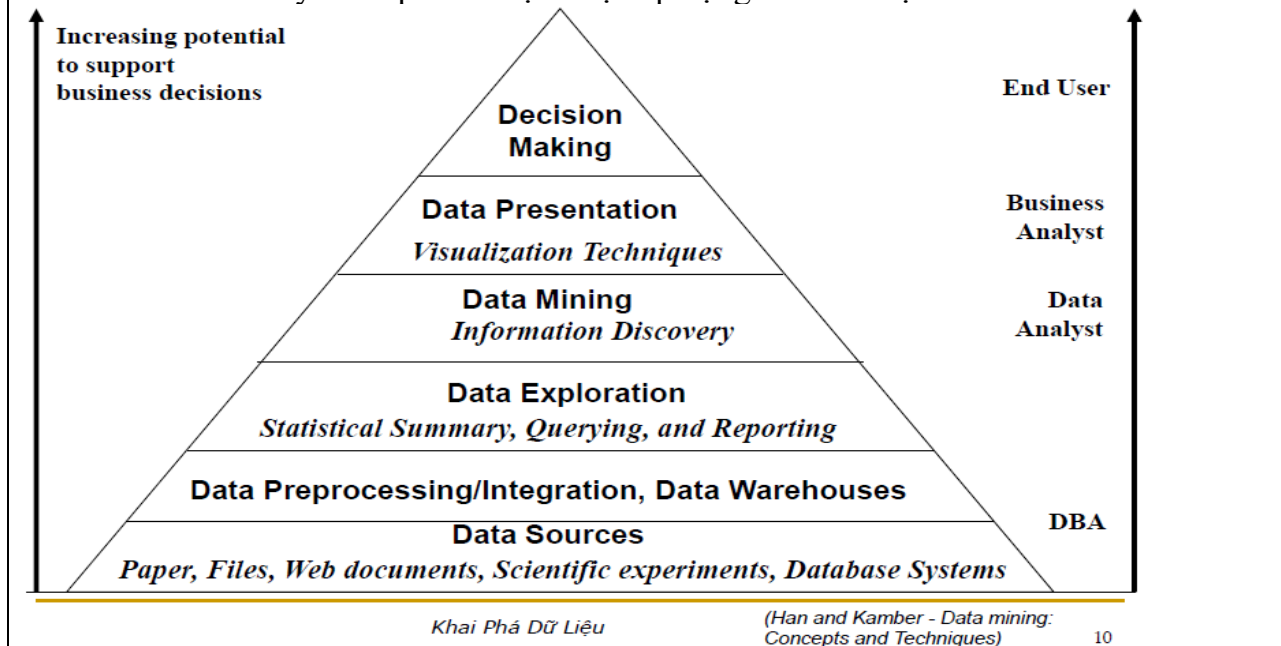
Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phản nội dung câu hỏi:

Biểu đồ sau cho thấy Khai phá dữ liệu được áp dụng cho lĩnh vực nào?



Các đáp án:

| | |
|---|------------|
| A | Chính trị |
| B | Kinh doanh |
| C | Giáo dục |
| D | Xã hội |

Đáp án đúng: B

Câu hỏi số: 09

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

Các lĩnh vực nào sau đây có liên quan đến kỹ thuật Khai phá dữ liệu?

Các đáp án:

| | |
|---|---------------------------------------------------|
| A | Công nghệ cơ sở dữ liệu (Database Technology) |
| B | Giải thuật (Algorithm) |
| C | Thống kê (Statistics), Học máy (Machine Learning) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 10

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

Các lĩnh vực nào sau đây có liên quan đến kỹ thuật Khai phá dữ liệu?

Các đáp án:

| | |
|---|------------------------------------------------------|
| A | Nhận dạng mẫu (Pattern recognition) |
| B | Hiển thị hóa (Visualization) |
| C | Tính toán hiệu năng cao (High-performance computing) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 11

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Hiểu

Mức độ: Dễ

Phần nội dung câu hỏi:

Các giải thuật phân tích dữ liệu cần phải hoạt động tốt với lượng dữ liệu rất lớn có thể lên đến bao nhiêu bytes?

Các đáp án:

| | |
|---|----------------|
| A | Gigabytes |
| B | Terabytes |
| C | Petabytes |
| D | Không giới hạn |

Đáp án đúng: C

Câu hỏi số: 12

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong khai phá dữ liệu, dữ liệu thường được biểu diễn trong không gian bao nhiêu chiều?

Các đáp án:

| | |
|---|---|
| A | 1 |
| B | 2 |

| | |
|---|-----------------|
| C | 3 |
| D | Không giới hạn. |

Đáp án đúng: D

Câu hỏi số: 13

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Phân tích

Mức độ: Trung bình

Phản nội dung câu hỏi:

Dữ liệu được khai phá thường có độ phức tạp (rất) cao. Vì sao?

Các đáp án:

| | |
|---|-----------------------------------------------------------------------|
| A | Các luồng dữ liệu và dữ liệu thu nhận từ các mạng cảm biến. |
| B | Dữ liệu phụ thuộc theo không gian/ thời gian, dữ liệu đa phương tiện. |
| C | Dữ liệu có cấu trúc, dữ liệu dạng đồ thị, các mạng xã hội. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 14

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Phân tích

Mức độ: Trung bình

Phản nội dung câu hỏi:

Khai phá dữ liệu thường có nhiều cách nhìn (quan điểm), tùy thuộc vào yếu tố nào sau đây?

Các đáp án:

| | |
|---|------------------------------------------------|
| A | Dữ liệu được khai phá; Tri thức được khám phá. |
| B | Các kỹ thuật được sử dụng. |
| C | Các ứng dụng (bài toán) thực tế. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 15

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phản nội dung câu hỏi:

Theo chức năng tổng quát, khai phá dữ liệu được chia thành mấy loại?

Các đáp án:

| | |
|---|---|
| A | 2 |
| B | 3 |
| C | 4 |
| D | 5 |

Đáp án đúng: B

Câu hỏi số: 16

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Phân tích

Mức độ: Trung bình

Phản nội dung câu hỏi:

Quá trình tích hợp thông tin và xây dựng các kho dữ liệu, trong khai phá dữ liệu được thực hiện theo qui trình nào sau đây?

Các đáp án:

| | |
|---|--------------------------------------------------------------------------------------------|
| A | Làm sạch dữ liệu -> Chuyển đổi dữ liệu -> Tích hợp dữ liệu -> Mô hình dữ liệu nhiều chiều. |
| B | Làm sạch dữ liệu -> Tích hợp dữ liệu -> Chuyển đổi dữ liệu -> Mô hình dữ liệu nhiều chiều. |
| C | Chuyển đổi dữ liệu -> Làm sạch dữ liệu -> Mô hình dữ liệu -> Tích hợp dữ liệu nhiều chiều. |
| D | Chuyển đổi dữ liệu -> Tích hợp dữ liệu -> Làm sạch dữ liệu -> Mô hình dữ liệu nhiều chiều. |

Đáp án đúng: A

Câu hỏi số: 17

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Biết

Mức độ: Khó

Phản nội dung câu hỏi:

Công nghệ khối dữ liệu (data cube) trong khai phá dữ liệu đề cập đến vấn đề gì?

Các đáp án:

| | |
|---|------------------------------------------------------------------------|
| A | Xử lý phân tích Offline. |
| B | Các phương pháp hiệu quả để tính toán kết hợp nhiều chiều của dữ liệu. |
| C | Tích hợp dữ liệu |
| D | Mô hình hóa dữ liệu. |

Đáp án đúng: B

Câu hỏi số: 18

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Phân tích

Mức độ: Trung bình

Phản nội dung câu hỏi:

Trong mô hình dữ liệu nhiều chiều (multi-dimensional data model), các khái niệm được mô tả theo nhiều chiều, nghĩa là gì?

Các đáp án:

| | |
|---|--------------------------------|
| A | Sự giống nhau và sự phân biệt. |
| B | Sự giống nhau và sự đặc trưng. |
| C | Sự đặc trưng và sự phân biệt. |
| D | Sự giống và khác nhau. |

Đáp án đúng: C

Câu hỏi số: 19

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phần nội dung câu hỏi:

Phương pháp nào sau đây thường dùng để khám phá các mẫu (luật) hoặc các tập mục (itemset) thường xuyên, trong các tập dữ liệu lớn?

Các đáp án:

| | |
|---|-----------------------------------|
| A | Phân tích kết hợp và tương quan. |
| B | Phân lớp và dự đoán. |
| C | Phân cụm và phân tích ngoại lai. |
| D | Phân tích xu hướng và tiến triển. |

Đáp án đúng: A

Câu hỏi số: 20

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phần nội dung câu hỏi:

Các phương pháp điển hình như Cây quyết định (Decision tree learning), Phân lớp Naïve Bayes, Máy véc-tơ hỗ trợ, Mạng nơ-ron nhân tạo (Artificial neural networks), Học quy nạp luật (Rule induction), Hồi quy tuyến tính (Lineal regression),... thường được sử dụng khi nào?

Các đáp án:

| | |
|---|-----------------------------------|
| A | Phân tích kết hợp và tương quan. |
| B | Phân lớp và dự đoán. |
| C | Phân cụm và phân tích ngoại lai. |
| D | Phân tích xu hướng và tiến triển. |

Đáp án đúng: B

Câu hỏi số: 21

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phần nội dung câu hỏi:

Các ứng dụng điển hình như phát hiện gian lận thẻ tín dụng, quảng cáo trực tuyến, phân loại/dự đoán các loại bệnh, phân loại các trang Web,... là kết quả của quá trình nào sau đây?

Các đáp án:

| | |
|---|-----------------------------------|
| A | Phân tích kết hợp và tương quan. |
| B | Phân lớp và dự đoán. |
| C | Phân cụm và phân tích ngoại lai. |
| D | Phân tích xu hướng và tiến triển. |

Đáp án đúng: A

Câu hỏi số: 22

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phần nội dung câu hỏi:

Phần việc nào sau đây được thực hiện trong quá trình Phân lớp (Classification) và Dự đoán (Prediction)?

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------|
| A | Xây dựng các mô hình (các hàm mục tiêu) dựa trên một số ví dụ học/huấn luyện. |
| B | Mô tả và phân biệt các lớp (các khái niệm) cho việc dự đoán trong tương lai. |
| C | Phân lớp các ví dụ mới, hoặc dự đoán các giá trị kiểu số. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 23

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

| |
|------------------------------------|
| Phân cụm (Cluster analysis) là gì? |
|------------------------------------|

Các đáp án:

| | |
|---|-----------------------------------------------------------------------------------------------------------------------------|
| A | Cực đại hóa sự tương tự giữa các đối tượng trong cùng một cụm – nhưng cực tiểu hóa sự tương tự giữa các đối tượng khác cụm. |
| B | Phương pháp học không giám sát (unsupervised learning) – không có thông tin về nhãn lớp. |
| C | Nhóm dữ liệu thành từng cụm (cluster). |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 24

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

| |
|------------------------------------------------------------------------------------|
| Phát biểu nào sau đây là đúng về phân tích ngoại lai (Outlier analysis/detection)? |
|------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|----------------------------------------------------------------------------|
| A | Ngoại lai: một đối tượng rất khác biệt với các đối tượng khác (trong cụm). |
| B | Là tên của một tổ chức quốc tế về mã hóa. |
| C | Mã khóa riêng |
| D | Là chính sách bảo mật. |

Đáp án đúng: A

Câu hỏi số: 25

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Hiểu

Mức độ: Khó

Phần nội dung câu hỏi:

| |
|--------------------------------------------------------------------------------------------------------|
| Phân tích chuỗi (sequence), xu hướng (trend), và tiến triển (evolution) bao gồm công việc nào sau đây? |
|--------------------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------------------------------------------------------------------|
| A | Phân tích xu hướng và sự dịch chuyển (khỏi xu hướng) và khai phá các mẫu kiểu chuỗi (sequential patterns). |
| B | Phân tích tính chu kỳ (Periodicity analysis), và phân tích chuỗi sự kiện liên tục theo thời gian (time-series) và chuỗi dữ liệu sinh học. |

| | |
|---|------------------------------------------------------------|
| C | Phân tích dựa trên sự tương tự (Similarity-based analysis) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 26

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

| |
|-----------------------------------------------------------------|
| Phân tích mạng và cấu trúc bao gồm những hoạt động nào sau đây? |
|-----------------------------------------------------------------|

Các đáp án:

| | |
|---|----------------------------------------------------------|
| A | Khai phá đồ thị dữ liệu (Graph mining). |
| B | Phân tích mạng thông tin (Information network analysis). |
| C | Khai phá Web (Web mining) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 27

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Phân tích

Mức độ: Trung bình

Phần nội dung câu hỏi:

| |
|-----------------------------------------------------------------------------------------------------------|
| Khai phá đồ thị dữ liệu (Graph mining) trong quá trình phân tích mạng và cấu trúc cần thực hiện những gì? |
|-----------------------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|----------------------------------------------------------------------------------------------------------------------------------------|
| A | Tìm ra các đồ thị con (các phần của đồ thị ban đầu), các cây (dữ liệu XML), các cấu trúc dữ liệu (dữ liệu Web)... thường xuyên xảy ra. |
| B | Phân tích các tác nhân (các đối tượng, các nút) và các mối quan hệ (các cạnh) trong các mạng hỗn tạp. |
| C | Khai phá các liên kết (Link mining). |
| D | Cực đại hóa sự tương tự giữa các đối tượng trong cùng một cụm – nhưng cực tiểu hóa sự tương tự giữa các đối tượng khác cụm. |

Đáp án đúng: A

Câu hỏi số: 28

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Phân tích

Mức độ: Trung bình

Phần nội dung câu hỏi:

| |
|--------------------------------------------------------------|
| Phát biểu nào sau đây là đúng về tầm quan trọng của các mẫu? |
|--------------------------------------------------------------|

Các đáp án:

| | |
|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A | Tất cả các mẫu đều quan trọng. |
| B | Một mẫu là quan trọng nếu nó dễ hiểu đối với người dùng, vẫn đúng đối với các dữ liệu mới, hữu dụng, mới mẻ, hoặc giúp xác nhận một giả thiết nào đó của một người dùng. |
| C | Một mẫu là quan trọng nếu nó khó hiểu đối với người sử dụng, hữu dụng, hoặc giúp xác nhận một giả thiết. |

| | |
|---|-------------------------------------------------------------------------------------------------------------|
| D | Đánh giá về tầm quan trọng của các mẫu phải dựa trên mục tiêu (Objective) và dựa trên chủ quan (Subjective) |
|---|-------------------------------------------------------------------------------------------------------------|

Đáp án đúng: B

Câu hỏi số: 29

Chương: Tổng quát về Khai phá dữ liệu

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

| |
|---------------------------------------------------------------------------|
| Đánh giá về mức độ quan trọng của mẫu, cần dựa trên tiêu chí nào sau đây? |
|---------------------------------------------------------------------------|

Các đáp án:

| | |
|---|----------------------------------------------------|
| A | Mức độ đơn giản (Simplicity). |
| B | Mức độ tin cậy (Certainty/Confidence). |
| C | Mức độ tiện ích (Utility) và tính mới mẻ (Novelty) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 30

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

| |
|-------------------------------------------------------------------------|
| Mức độ đơn giản trong đánh giá tầm quan trọng của mẫu bao gồm những gì? |
|-------------------------------------------------------------------------|

Các đáp án:

| | |
|---|---------------------------------------------------------------------------|
| A | Độ tin cậy của các luật kết hợp và Độ chính xác của phân lớp học được. |
| B | Độ dài của các luật kết hợp và Kích thước của cây quyết định học được. |
| C | Độ hỗ trợ của các luật kết hợp và Ngưỡng nhiễu đối với phân lớp học được. |
| D | Mẫu mới, chưa bao giờ được biết đến. |

Đáp án đúng: B

Câu hỏi số: 31

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

| |
|--------------------------------------------------------------------------------------------------|
| Mức độ tin cậy (Certainty/Confidence) trong đánh giá mức độ quan trọng của mẫu bao gồm những gì? |
|--------------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|---------------------------------------------------------------------------|
| A | Độ tin cậy của các luật kết hợp và Độ chính xác của phân lớp học được. |
| B | Độ dài của các luật kết hợp và Kích thước của cây quyết định học được. |
| C | Độ hỗ trợ của các luật kết hợp và Ngưỡng nhiễu đối với phân lớp học được. |
| D | Mẫu mới, chưa bao giờ được biết đến. |

Đáp án đúng: A

Câu hỏi số: 32

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Hiểu

Mức độ: Trung bình

Phân nội dung câu hỏi:

Mức độ tiện ích (Utility) là khả năng hữu ích của mẫu trong đánh giá mức độ quan trọng của mẫu bao gồm những gì?

Các đáp án:

| | |
|---|---------------------------------------------------------------------------|
| A | Độ tin cậy của các luật kết hợp và Độ chính xác của phân lớp học được. |
| B | Độ dài của các luật kết hợp và Kích thước của cây quyết định học được. |
| C | Độ hỗ trợ của các luật kết hợp và Ngưỡng nhiễu đối với phân lớp học được. |
| D | Mẫu mới, chưa bao giờ được biết đến. |

Đáp án đúng: C

Câu hỏi số: 33

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Hiểu

Mức độ: Trung bình

Phân nội dung câu hỏi:

Khi hiển thị các mẫu tìm được, các phát biểu nào sau đây là đúng?

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------------------------------------------|
| A | Người dùng khác nhau, mục đích sử dụng khác nhau sẽ yêu cầu các dạng hiển thị khác nhau đối với các mẫu tìm được. |
| B | Sự phân cấp khái niệm cho phép nhìn (xét) dữ liệu theo các cách nhìn khác nhau. |
| C | Các kiểu tri thức khác nhau đòi hỏi các cách biểu diễn khác nhau (đối với các mẫu tìm được). |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 34

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Áp dụng

Mức độ: Dễ

Phân nội dung câu hỏi:

Các ứng dụng tiềm năng của Khai phá dữ liệu bao gồm?

Các đáp án:

| | |
|---|---------------------------------------------------------------|
| A | Phân tích và quản lý thị trường. |
| B | Phân tích và quản lý rủi ro. |
| C | Phát hiện gian lận và phát hiện các mẫu bất thường (Outliers) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 35

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Áp dụng

Mức độ: Dễ

Phân nội dung câu hỏi:

Các ứng dụng tiềm năng của Khai phá dữ liệu trong lĩnh vực Phân tích thị trường bao gồm?

Các đáp án:

| | |
|---|-----------------------------------------------|
| A | Quảng cáo cá nhân (Target marketing). |
| B | Phân tích thị trường (Cross-market analysis). |
| C | Lập hồ sơ khách hàng (Customer profiling). |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 36

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Áp dụng

Mức độ: Dễ

Phân nội dung câu hỏi:

| |
|-------------------------------------------------------------------------------------|
| Các ứng dụng tiềm năng của Khai phá dữ liệu trong lĩnh vực Quản lý rủi ro, bao gồm? |
|-------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|---------------------------------------------|
| A | Lập kế hoạch tài chính và đánh giá tài sản. |
| B | Lập kế hoạch sử dụng tài nguyên. |
| C | Cạnh tranh trong kinh doanh. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 37

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Áp dụng

Mức độ: Dễ

Phân nội dung câu hỏi:

| |
|-----------------------------------------------------------------------------------------|
| Các ứng dụng tiềm năng của Khai phá dữ liệu trong lĩnh vực Phát hiện gian lận, bao gồm? |
|-----------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------|
| A | Phân cụm và xây dựng mô hình dự đoán gian lận, phân tích ngoại lai (outlier). |
| B | Các ứng dụng chăm sóc sức khỏe như bảo hiểm y tế. |
| C | Các ứng dụng trong Viễn thông, các dịch vụ sử dụng thẻ tín dụng. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 38

Chương: Tổng quan về Khai phá dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phân nội dung câu hỏi:

| |
|-----------------------------------------------------|
| Các vấn đề thách thức trong Khai phá dữ liệu là gì? |
|-----------------------------------------------------|

Các đáp án:

| | |
|---|-----------------------------------------------------------------------------------------------------------------|
| A | Tính hiệu quả (efficiency) và tính ổn định (scalability) của các giải thuật khai phá dữ liệu. |
| B | Xử lý dữ liệu có số chiều (số thuộc tính) lớn. Xử lý với dữ liệu chứa nhiễu, không chắc chắn, không hoàn chỉnh. |
| C | Bảo đảm tính an ninh, toàn vẹn, riêng tư trong khai phá dữ liệu. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 39

Chương: Giới thiệu về WEKA-Waikato Environment for Knowledge Analysis

Kỹ năng: Biết

Mức độ: Trung bình

Phân nội dung câu hỏi:

Phát biểu nào sau đây về công cụ WEKA là đúng?

Các đáp án:

| | |
|---|------------------------------------------------------------------------------------|
| A | WEKA là một phần mềm viết bằng Java, phục vụ lĩnh vực học máy và khai phá dữ liệu. |
| B | WEKA là một công cụ viết bằng PHP, phục vụ lĩnh vực khai phá dữ liệu. |
| C | WEKA là một công cụ phần mềm viết bằng Visual Basic, phục vụ lĩnh vực học máy. |
| D | WEKA là một phần mềm viết bằng R, phục vụ lĩnh vực khai phá dữ liệu. |

Đáp án đúng: A

Câu hỏi số: 40

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Trung bình

Phân nội dung câu hỏi:

Các tính năng chính của Phần mềm WEKA gồm có?

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------------------------------------------------|
| A | Một tập các công cụ tiền xử lý dữ liệu, các giải thuật học máy, khai phá dữ liệu, và các phương pháp thí nghiệm đánh giá. |
| B | Giao diện đồ họa (gồm các tính năng hiển thị hóa dữ liệu). |
| C | Môi trường cho phép so sánh các giải thuật học máy và khai phá dữ liệu. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 41

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Trung bình

Phân nội dung câu hỏi:

Các môi trường chính của WEKA gồm có?

Các đáp án:

| | |
|---|----------------------------------------------------------------------------------------|
| A | Simple CLI là giao diện đơn giản kiểu dòng lệnh (như MS-DOS). |
| B | Explorer môi trường cho phép sử dụng tất cả các khả năng của WEKA để khám phá dữ liệu. |
| C | Experimenter và KnowledgeFlow. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 42

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

Môi trường nào trong WEKA cho phép sử dụng tất cả các khả năng của WEKA để khám phá dữ liệu?

Các đáp án:

| | |
|---|---------------|
| A | Explorer |
| B | Experimenter |
| C | Simple CLI |
| D | KnowledgeFlow |

Đáp án đúng: A

Câu hỏi số: 43

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

Môi trường nào trong WEKA cho phép bạn tương tác đồ họa kiểu kéo/thả để thiết kế các bước (các thành phần) của một thí nghiệm?

Các đáp án:

| | |
|---|---------------|
| A | Explorer |
| B | Experimenter |
| C | Simple CLI |
| D | KnowledgeFlow |

Đáp án đúng: D

Câu hỏi số: 44

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

Môi trường nào trong WEKA cho phép tiến hành các thí nghiệm và thực hiện các kiểm tra thống kê (statistical tests) giữa các mô hình học máy?

Các đáp án:

| | |
|---|---------------|
| A | Explorer |
| B | Experimenter |
| C | Simple CLI |
| D | KnowledgeFlow |

Đáp án đúng: B

Câu hỏi số: 45

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Dễ

Phần nội dung câu hỏi:

Môi trường nào trong WEKA là giao diện đơn giản kiểu dòng lệnh như MS-DOS?

Các đáp án:

| | |
|---|---------------|
| A | Explorer |
| B | Experimenter |
| C | Simple CLI |
| D | KnowledgeFlow |

Đáp án đúng: C

Câu hỏi số: 46

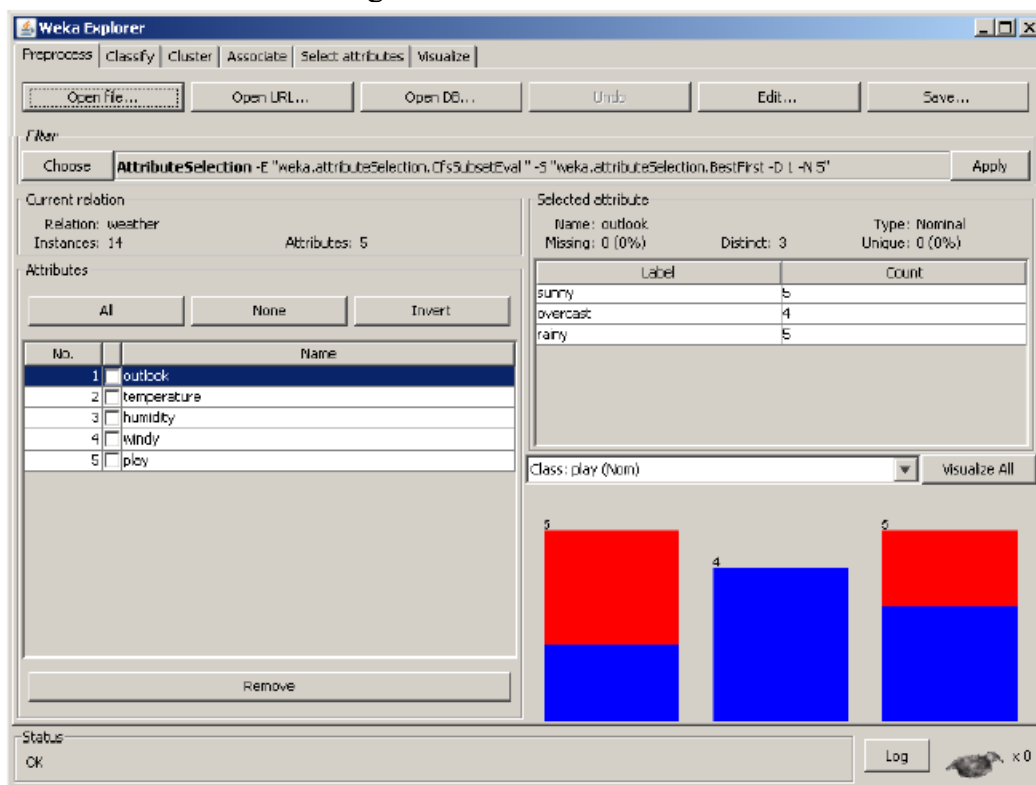
Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Khó

Phản nội dung câu hỏi:

Giao diện sau là môi trường làm việc nào của WEKA?



Khai Phá Dữ Liệu

Các đáp án:

| | |
|---|---------------|
| A | Explorer |
| B | Experimenter |
| C | Simple CLI |
| D | KnowledgeFlow |

Đáp án đúng: A

Câu hỏi số: 47

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Khó

Phản nội dung câu hỏi:

Trong môi trường Explorer của WEKA, để chọn và thay đổi dữ liệu làm việc, tính năng nào sau đây sẽ được chọn?

Các đáp án:

| | |
|---|------------|
| A | Preprocess |
| B | Classify |
| C | Cluster |
| D | Associate |

Đáp án đúng: A

Câu hỏi số: 48

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Khó

Phân nội dung câu hỏi:

Trong môi trường Explorer của WEKA, để huấn luyện và kiểm tra các mô hình học máy, tính năng nào sau đây sẽ được chọn?

Các đáp án:

| | |
|---|------------|
| A | Preprocess |
| B | Classify |
| C | Cluster |
| D | Associate |

Đáp án đúng: B

Câu hỏi số: 49

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phân nội dung câu hỏi:

Khái niệm Tập dữ liệu (Dataset) là gì?

Các đáp án:

| | |
|---|-----------------------------------------------------------------------------------------------------------|
| A | Một tập dữ liệu (Dataset) là một tập các đối tượng (Objects) được lựa chọn theo các thuộc tính cho trước. |
| B | Một tập dữ liệu (Dataset) là một tập các thuộc tính của đối tượng (Objects). |
| C | Một tập dữ liệu (Dataset) là một tập hợp các đối tượng (Objects) và các phương tiện lưu trữ chúng. |
| D | Một tập dữ liệu (Dataset) là một tập hợp các đối tượng (Objects) và các thuộc tính của chúng. |

Đáp án đúng: D

Câu hỏi số: 50

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Dễ

Phân nội dung câu hỏi:

Trong một tập dữ liệu, mỗi thuộc tính (attribute) mô tả bao nhiêu đặc điểm của một đối tượng?

Các đáp án:

| | |
|---|---|
| A | 1 |
| B | 2 |

| | |
|---|---|
| C | 3 |
| D | 4 |

Đáp án đúng: A

Câu hỏi số: 51

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Dễ

Phản nội dung câu hỏi:

Trong một tập dữ liệu, mỗi thuộc tính (attribute) mô tả một đặc điểm của bao nhiêu đối tượng?

Các đáp án:

| | |
|---|---|
| A | 1 |
| B | 4 |
| C | 3 |
| D | 2 |

Đáp án đúng: A

Câu hỏi số: 52

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Dễ

Phản nội dung câu hỏi:

Trong tiền xử lý dữ liệu, một tập các giá trị của các thuộc tính sẽ mô tả bao nhiêu đối tượng?

Các đáp án:

| | |
|---|---|
| A | 2 |
| B | 4 |
| C | 1 |
| D | 3 |

Đáp án đúng: C

Câu hỏi số: 53

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Dễ

Phản nội dung câu hỏi:

Trong tập dữ liệu, khái niệm “đối tượng” còn được tham chiếu đến với các tên gọi khác là gì?

Các đáp án:

| | |
|---|----------------------------|
| A | Bản ghi (record) |
| B | Trường hợp (case) |
| C | Mẫu (sample) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 54

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

Có các kiểu tập dữ liệu nào?

Các đáp án:

| | |
|---|----------------------------|
| A | Bản ghi (record) |
| B | Đồ thị (graph) |
| C | Có trật tự (ordered) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 55

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Khó

Phần nội dung câu hỏi:

Trong CSDL quan hệ, các tập dữ liệu thường được biểu diễn theo kiểu nào?

Các đáp án:

| | |
|---|----------------------------|
| A | Bản ghi (record) |
| B | Đồ thị (graph) |
| C | Có trật tự (ordered) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: A

Câu hỏi số: 56

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Khó

Phần nội dung câu hỏi:

Tập ma trận dữ liệu thường được biểu diễn theo kiểu nào?

Các đáp án:

| | |
|---|----------------------------|
| A | Bản ghi (record) |
| B | Đồ thị (graph) |
| C | Có trật tự (ordered) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: A

Câu hỏi số: 57

Chương: Tiền xử lý dữ liệu

Kỹ năng: Phân tích

Mức độ: Dễ

Phần nội dung câu hỏi:

Tập dữ liệu văn bản (document) thường được biểu diễn theo kiểu nào?

Các đáp án:

| | |
|---|------------------|
| A | Bản ghi (record) |
|---|------------------|

| | |
|---|----------------------------|
| B | Đồ thị (graph) |
| C | Có trật tự (ordered) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: A

Câu hỏi số: 58

Chương: Tiền xử lý dữ liệu

Kỹ năng: Phân tích

Mức độ: Dễ

Phân nội dung câu hỏi:

Dữ liệu giao dịch thường được biểu diễn theo kiểu nào?

Các đáp án:

| | |
|---|----------------------------|
| A | Bản ghi (record) |
| B | Đồ thị (graph) |
| C | Có trật tự (ordered) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: A

Câu hỏi số: 59

Chương: Tiền xử lý dữ liệu

Kỹ năng: Phân tích

Mức độ: Trung bình

Phân nội dung câu hỏi:

Tiền xử lý dữ liệu từ mạng thông tin, mạng xã hội, hoặc World Wide Web thường được biểu diễn theo kiểu nào?

Các đáp án:

| | |
|---|----------------------------|
| A | Bản ghi (record) |
| B | Đồ thị (graph) |
| C | Có trật tự (ordered) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: B

Câu hỏi số: 60

Chương: Tiền xử lý dữ liệu

Kỹ năng: Phân tích

Mức độ: Trung bình

Phân nội dung câu hỏi:

Dữ liệu có cấu trúc phân tử (Molecular structures) thường được biểu diễn theo kiểu nào trong quá trình tiền xử lý?

Các đáp án:

| | |
|---|----------------------------|
| A | Bản ghi (record) |
| B | Đồ thị (graph) |
| C | Có trật tự (ordered) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: B

Câu hỏi số: 61

Chương: Tiền xử lý dữ liệu

Kỹ năng: Phân tích

Mức độ: Khó

Phản nội dung câu hỏi:

Dữ liệu không gian (vd: bản đồ) thường được biểu diễn theo kiểu tập dữ liệu nào trong quá trình tiền xử lý?

Các đáp án:

| | |
|---|----------------------------|
| A | Bản ghi (record) |
| B | Đồ thị (graph) |
| C | Có trật tự (ordered) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: C

Câu hỏi số: 62

Chương: Tiền xử lý dữ liệu

Kỹ năng: Phân tích

Mức độ: Khó

Phản nội dung câu hỏi:

Dữ liệu thời gian (vd: time-series data) thường được biểu diễn theo kiểu tập dữ liệu nào trong quá trình tiền xử lý?

Các đáp án:

| | |
|---|----------------------------|
| A | Bản ghi (record) |
| B | Đồ thị (graph) |
| C | Có trật tự (ordered) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: C

Câu hỏi số: 63

Chương: Tiền xử lý dữ liệu

Kỹ năng: Phân tích

Mức độ: Khó

Phản nội dung câu hỏi:

Dữ liệu chuỗi như chuỗi giao dịch, thường được biểu diễn theo kiểu tập dữ liệu nào trong quá trình tiền xử lý?

Các đáp án:

| | |
|---|----------------------------|
| A | Bản ghi (record) |
| B | Đồ thị (graph) |
| C | Có trật tự (ordered) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: C

Câu hỏi số: 64

Chương: Tiền xử lý dữ liệu

Kỹ năng: Phân tích

Mức độ: Khó

Phản nội dung câu hỏi:

Dữ liệu chuỗi di truyền (genetic sequence data) thường được biểu diễn theo kiểu tập dữ liệu nào trong quá trình tiền xử lý?

Các đáp án:

| | |
|---|----------------------------|
| A | Bản ghi (record) |
| B | Đồ thị (graph) |
| C | Có trật tự (ordered) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: C

Câu hỏi số: 65

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Khó

Phần nội dung câu hỏi:

| |
|--------------------------------------------------------------|
| Trong tiền xử lý dữ liệu có các kiểu giá trị thuộc tính nào? |
|--------------------------------------------------------------|

Các đáp án:

| | |
|---|--------------------------------|
| A | Kiểu định danh/chuỗi (nominal) |
| B | Kiểu nhị phân (binary) |
| C | Kiểu có thứ tự (ordinal) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 66

Chương: Tiền xử lý dữ liệu

Kỹ năng: Hiểu

Mức độ: Dễ

Phần nội dung câu hỏi:

| |
|--------------------------------------------------------------------------------------------------|
| Kiểu giá trị thuộc tính nào mà trong đó giá trị được lấy từ một tập không có thứ tự các giá trị? |
|--------------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|--------------------------------|
| A | Kiểu định danh/chuỗi (nominal) |
| B | Kiểu nhị phân (binary) |
| C | Kiểu có thứ tự (ordinal) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: A

Câu hỏi số: 67

Chương: Tiền xử lý dữ liệu

Kỹ năng: Hiểu

Mức độ: Khó

Phần nội dung câu hỏi:

| |
|-------------------------------------------------------------------------------------------|
| Kiểu giá trị thuộc tính nào trong đó tập các giá trị chỉ gồm hai giá trị (Y/N, 0/1, T/F)? |
|-------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|--------------------------------|
| A | Kiểu định danh/chuỗi (nominal) |
| B | Kiểu nhị phân (binary) |
| C | Kiểu có thứ tự (ordinal) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: B

Câu hỏi số: 68

Chương: Tiền xử lý dữ liệu

Kỹ năng: Hiểu

Mức độ: Khó

Phản nội dung câu hỏi:

Kiểu giá trị thuộc tính nào sau đây lấy giá trị từ một tập có thứ tự các giá trị như các thuộc tính lấy giá trị số (age, height,...)?

Các đáp án:

| | |
|---|--------------------------------|
| A | Kiểu định danh/chuỗi (nominal) |
| B | Kiểu nhị phân (binary) |
| C | Kiểu có thứ tự (ordinal) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: B

Câu hỏi số: 69

Chương: Tiền xử lý dữ liệu

Kỹ năng: Phân tích

Mức độ: Dễ

Phản nội dung câu hỏi:

Kiểu thuộc tính rời rạc (Discrete-valued attributes) trong tiền xử lý dữ liệu có đặc điểm gì?

Các đáp án:

| | |
|---|-------------------------------------------------------------|
| A | Tập các giá trị là một tập hữu hạn. |
| B | Bao gồm cả các thuộc tính có kiểu giá trị là các số nguyên. |
| C | Bao gồm cả các thuộc tính nhị phân (binary attributes). |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 70

Chương: Tiền xử lý dữ liệu

Kỹ năng: Phân tích

Mức độ: Trung bình

Phản nội dung câu hỏi:

Kiểu thuộc tính liên tục (Continuous-valued attributes) của dữ liệu có đặc điểm gì?

Các đáp án:

| | |
|---|-------------------------------------------------------------|
| A | Tập các giá trị là một tập hữu hạn. |
| B | Bao gồm cả các thuộc tính có kiểu giá trị là các số nguyên. |
| C | Bao gồm cả các thuộc tính nhị phân (binary attributes). |
| D | Các giá trị là các số thực (real numbers). |

Đáp án đúng: D

Câu hỏi số: 71

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phản nội dung câu hỏi:

Các đặc tính mô tả dữ liệu nhằm mục đích gì?

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A | Để hiểu rõ về sự phân bố của dữ liệu như giá trị cực tiểu/ cực đại, giá trị xuất hiện nhiều nhất, giá trị trung bình, giá trị trung vị, sự biến thiên và độ lệch chuẩn, và các ngoại lai. |
| B | Để hiểu rõ về dữ liệu có được như chiều hướng chính/trung tâm, sự biến thiên, sự phân bố. |
| C | Để biểu diễn dữ liệu bằng các phương pháp hiển thị đồ họa, giúp hiểu rõ các đặc điểm của dữ liệu. |
| D | Cung cấp cái nhìn định tính đối với các tập dữ liệu lớn. |

Đáp án đúng: B

Câu hỏi số: 72

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

| |
|-----------------------------------------------------------|
| Sự phân bố của dữ liệu (data dispersion) gồm có những gì? |
|-----------------------------------------------------------|

Các đáp án:

| | |
|---|---------------------------------------------------------|
| A | Giá trị cực tiểu/cực đại; Giá trị xuất hiện nhiều nhất. |
| B | Giá trị trung bình; Giá trị trung vị. |
| C | Sự biến thiên, độ lệch chuẩn, và Các ngoại lai. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 73

Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phần nội dung câu hỏi:

| |
|------------------------------------------------------------------------------------------|
| Tác dụng của hiển thị hóa dữ liệu (data visualization) trong quá trình tiền xử lý là gì? |
|------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|------------------------------------------------------------------------------------------------------------|
| A | Cung cấp cái nhìn định tính đối với các tập dữ liệu lớn. |
| B | Có thể chỉ ra các mẫu, các xu hướng, các cấu trúc, các bất thường, và các quan hệ trong dữ liệu. |
| C | Hỗ trợ xác định các vùng dữ liệu quan trọng và các tham số phù hợp cho các phân tích định lượng tiếp theo. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 74

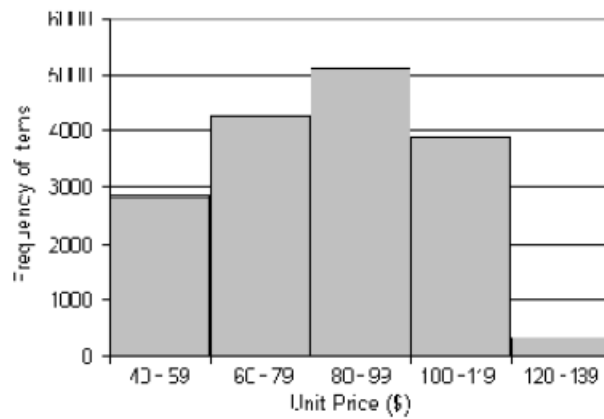
Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Dễ

Phần nội dung câu hỏi:

| |
|-----------------------------------------------------|
| Phát biểu nào sau đây là đúng về Biểu đồ histogram? |
|-----------------------------------------------------|



(Han, Kamber - Data Mining: Concepts and Techniques)

Các đáp án:

| | |
|---|-------------------------------------------------------------------|
| A | Biểu đồ histogram là cách biểu diễn dựa trên đồ thị. |
| B | Được sử dụng rất phổ biến. |
| C | Hiện thị các mô tả thống kê xuất hiện theo một thuộc tính nào đó. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 75

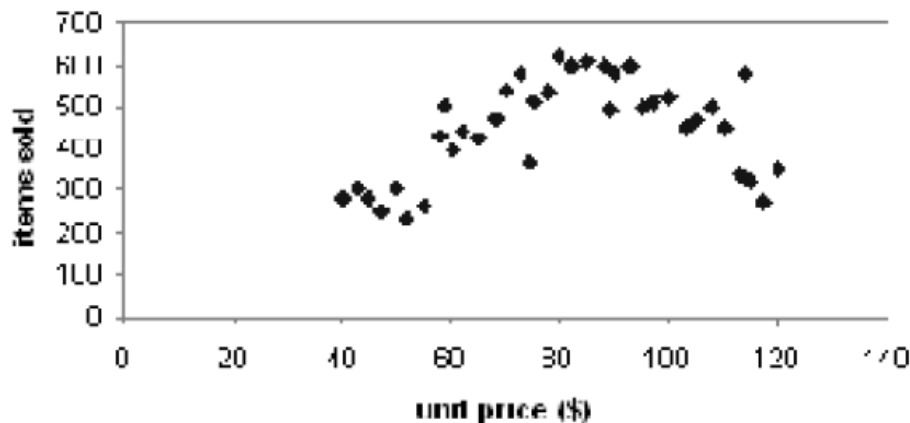
Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Khó

Phần nội dung câu hỏi:

Phát biểu nào sau đây là đúng về Đồ thị rải rác (Scatter plot)?



(Han, Kamber - Data Mining: Concepts and Techniques)

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------------------------------|
| A | Cho phép hiển thị quan hệ 2 chiều (giữa 2 thuộc tính) của dữ liệu. |
| B | Cho phép quan sát (trực quan) các nhóm điểm, các ngoại lai. |
| C | Mỗi cặp giá trị của 2 thuộc tính được xét tương ứng với 2 tọa độ của điểm được hiển thị trên mặt phẳng. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 76

Chương: Tiền xử lý dữ liệu

Kỹ năng: Áp dụng

Mức độ: Khó

Phân nội dung câu hỏi:

Các nhiệm vụ chính của tiền xử lý dữ liệu là gì?

Các đáp án:

| | |
|---|----------------------------------------|
| A | Làm sạch dữ liệu (data cleaning) |
| B | Tích hợp dữ liệu (data integration) |
| C | Biến đổi dữ liệu (data transformation) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 77

Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phân nội dung câu hỏi:

Gán các giá trị thuộc tính còn thiếu, sửa chữa các dữ liệu nhiễu/lỗi, xác định hoặc loại bỏ các ngoại lai, giải quyết các mâu thuẫn dữ liệu, là những thao tác của quá trình nào?

Các đáp án:

| | |
|---|------------------|
| A | Làm sạch dữ liệu |
| B | Tích hợp dữ liệu |
| C | Biến đổi dữ liệu |
| D | Giảm bớt dữ liệu |

Đáp án đúng: A

Câu hỏi số: 78

Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Khó

Phân nội dung câu hỏi:

Tích hợp nhiều cơ sở dữ liệu, nhiều khối dữ liệu (data cubes), hoặc nhiều tập tin dữ liệu, là các thao tác của quá trình nào?

Các đáp án:

| | |
|---|------------------|
| A | Làm sạch dữ liệu |
| B | Tích hợp dữ liệu |
| C | Biến đổi dữ liệu |
| D | Giảm bớt dữ liệu |

Đáp án đúng: D

Câu hỏi số: 79

Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Khó

Phân nội dung câu hỏi:

Chuẩn hóa (normalize) và kết hợp (aggregate) dữ liệu, là 2 thao tác của quá trình nào?

Các đáp án:

| | |
|---|------------------|
| A | Làm sạch dữ liệu |
| B | Tích hợp dữ liệu |
| C | Biến đổi dữ liệu |
| D | Giảm bớt dữ liệu |

Đáp án đúng: C

Câu hỏi số: 80

Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phản nội dung câu hỏi:

Giảm bớt về biểu diễn (các thuộc tính) của dữ liệu, giảm bớt kích thước dữ liệu, nhưng vẫn đảm bảo thu được các kết quả khai phá dữ liệu tương đương, là các thao tác của quá trình nào?

Các đáp án:

| | |
|---|------------------|
| A | Làm sạch dữ liệu |
| B | Tích hợp dữ liệu |
| C | Biến đổi dữ liệu |
| D | Giảm bớt dữ liệu |

Đáp án đúng: D

Câu hỏi số: 81

Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phản nội dung câu hỏi:

Rời rạc hóa dữ liệu (Data discretization) được sử dụng đối với các dữ liệu có các thuộc tính kiểu số, là một thao tác trong quá trình nào?

Các đáp án:

| | |
|---|------------------|
| A | Làm sạch dữ liệu |
| B | Tích hợp dữ liệu |
| C | Biến đổi dữ liệu |
| D | Giảm bớt dữ liệu |

Đáp án đúng: D

Câu hỏi số: 82

Chương: Tiền xử lý dữ liệu

Kỹ năng: Phân tích

Mức độ: Dễ

Phản nội dung câu hỏi:

Nguồn gốc/lý do của dữ liệu không sạch là gì?

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------|
| A | Dữ liệu không hoàn chỉnh (incomplete): thiếu các thuộc tính hoặc thiếu một số |
|---|-------------------------------------------------------------------------------|

| | |
|---|----------------------------------------------------------------------------------------------|
| | thuộc tính. |
| B | Nhiều/lỗi (noise/error): chứa đựng những lỗi hoặc các ví dụ bất thường (abnormal instances). |
| C | Mâu thuẫn (inconsistent): chứa đựng các mâu thuẫn (không thống nhất) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 83

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

Nguyên nhân nào dẫn đến dữ liệu bị nhiều/lỗi?

Các đáp án:

| | |
|---|----------------------------|
| A | Do việc thu thập dữ liệu |
| B | Do việc nhập dữ liệu |
| C | Do việc truyền dữ liệu |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 84

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Dễ

Phần nội dung câu hỏi:

Giá trị của thuộc tính không có (not available) tại thời điểm được thu thập do các vấn đề gây ra bởi phần cứng, phần mềm, hoặc người thu thập phản ánh tính chất nào của dữ liệu khi chưa được làm sạch?

Các đáp án:

| | |
|---|----------------------------|
| A | Không hoàn chỉnh |
| B | Nhiều/lỗi |
| C | Mâu thuẫn |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: A

Câu hỏi số: 85

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

Dữ liệu được thu thập từ nhiều nguồn khác nhau, và vi phạm các ràng buộc (điều kiện) đối với các thuộc tính, phản ánh tính chất gì của dữ liệu trước khi được làm sạch?

Các đáp án:

| | |
|---|----------------------------|
| A | Không hoàn chỉnh |
| B | Nhiều/lỗi |
| C | Mâu thuẫn |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: C

Câu hỏi số: 86

Chương: Tiền xử lý dữ liệu

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

Tại sao cần phải làm sạch dữ liệu?

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------------------------------|
| A | Nếu dữ liệu không sạch thì các kết quả khai phá dữ liệu sẽ bị ảnh hưởng và không đáng tin cậy. |
| B | Các kết quả khai phá dữ liệu không chính xác sẽ dẫn đến các quyết định không chính xác, không tối ưu. |
| C | Dữ liệu thu được từ thực tế có thể chứa nhiều lỗi, nhiễu, không hoàn chỉnh, và có mâu thuẫn. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 87

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

Nguyên nhân nào dẫn đến thuộc tính của dữ liệu có thể bị thiếu?

Các đáp án:

| | |
|---|-----------------------------------------------------------------------------------|
| A | Lỗi của các thiết bị phần cứng. |
| B | Không tương thích với các dữ liệu đã được ghi từ trước, do đó giá trị mới bị xóa. |
| C | Dữ liệu không được nhập vào (lỗi của người nhập liệu). |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 88

Chương: Tiền xử lý dữ liệu

Kỹ năng: Áp dụng

Mức độ: Dễ

Phân nội dung câu hỏi:

Phân khoảng (Binning) bằng cách sắp xếp dữ liệu, và phân chia thành các khoảng (bins) có tần số xuất hiện giá trị như nhau là giải pháp được áp dụng khi nào?

Các đáp án:

| | |
|---|---------------------------|
| A | Thuộc tính thiếu giá trị. |
| B | Dữ liệu chứa nhiễu. |
| C | Dữ liệu có mâu thuẫn. |
| D | Dữ liệu không hoàn chỉnh. |

Đáp án đúng: B

Câu hỏi số: 89

Chương: Tiền xử lý dữ liệu

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phân nội dung câu hỏi:

Gắn dữ liệu với một hàm Hồi qui (Regression function) là giải pháp được sử dụng trong tiền xử lý dữ liệu khi nào?

Các đáp án:

| | |
|---|---------------------------|
| A | Thuộc tính thiếu giá trị. |
| B | Dữ liệu chứa nhiễu. |
| C | Dữ liệu có mâu thuẫn. |
| D | Dữ liệu không hoàn chỉnh. |

Đáp án đúng: B

Câu hỏi số: 90

Chương: Tiền xử lý dữ liệu

Kỹ năng: Áp dụng

Mức độ: Khó

Phân nội dung câu hỏi:

Giải pháp phân cụm (Clustering) phát hiện và loại bỏ các ngoại lai (sau khi đã xác định các cụm) được sử dụng trong tiền xử lý dữ liệu khi nào?

Các đáp án:

| | |
|---|---------------------------|
| A | Thuộc tính thiếu giá trị. |
| B | Dữ liệu chứa nhiễu. |
| C | Dữ liệu có mâu thuẫn. |
| D | Dữ liệu không hoàn chỉnh. |

Đáp án đúng: B

Câu hỏi số: 91

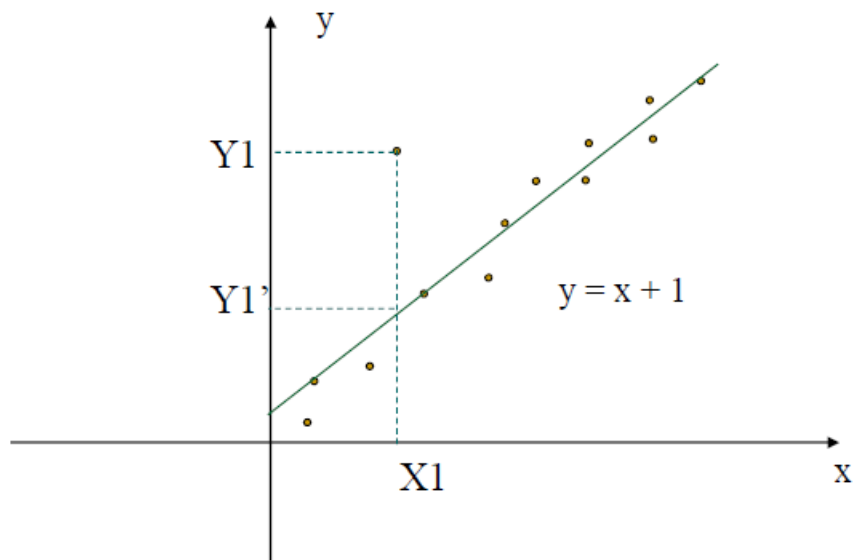
Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Khó

Phân nội dung câu hỏi:

Đồ thị sau thể hiện giải pháp gì?



(Han, Kamber - Data Mining: Concepts and Techniques)

Các đáp án:

| | |
|---|--------------------------------------|
| A | Hồi quy (Regression) |
| B | Phân khoảng (Binning) |
| C | Phân tích các cụm (Cluster Analysis) |
| D | Tích hợp dữ liệu |

Đáp án đúng: A

Câu hỏi số: 92

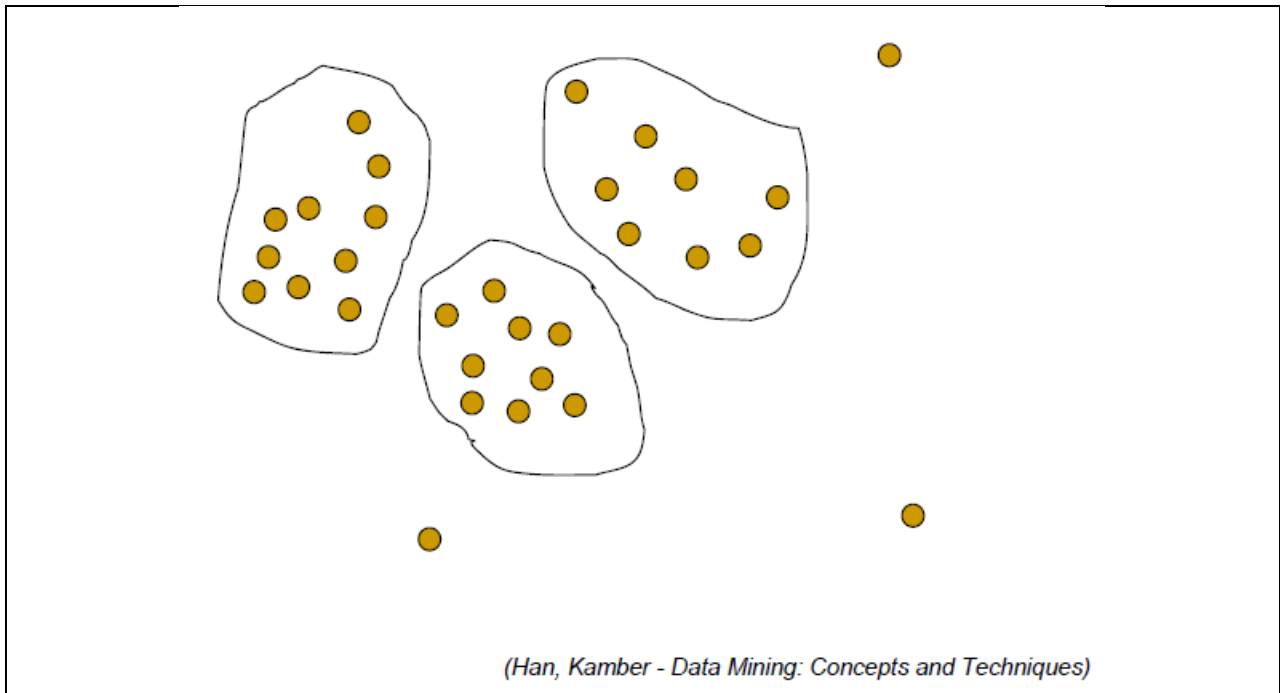
Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Dễ

Phân nội dung câu hỏi:

Đồ thị sau thể hiện giải pháp gì?



Các đáp án:

| | |
|---|--------------------------------------|
| A | Hồi quy (Regression) |
| B | Phân khoảng (Binning) |
| C | Phân tích các cụm (Cluster Analysis) |
| D | Tích hợp dữ liệu |

Đáp án đúng: C

Câu hỏi số: 93

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Trung bình

Phân nội dung câu hỏi:

Giải pháp phân chia dữ liệu thành các khoảng với độ sâu bằng nhau, hoặc biểu diễn khoảng dữ liệu bởi giá trị trung bình, còn được gọi là gì?

Các đáp án:

| | |
|---|--------------------------------------|
| A | Hồi quy (Regression) |
| B | Phân khoảng (Binning) |
| C | Phân tích các cụm (Cluster Analysis) |
| D | Tích hợp dữ liệu |

Đáp án đúng: B

Câu hỏi số: 94

Chương: Tiền xử lý dữ liệu

Kỹ năng: Hiểu

Mức độ: Trung bình

Phân nội dung câu hỏi:

Việc kết hợp dữ liệu từ nhiều nguồn vào một kho dữ liệu thống nhất còn được gọi là giải pháp gì?

Các đáp án:

| | |
|---|--------------------------------------|
| A | Hồi quy (Regression) |
| B | Phân khoảng (Binning) |
| C | Phân tích các cụm (Cluster Analysis) |
| D | Tích hợp dữ liệu |

Đáp án đúng: D

Câu hỏi số: 95

Chương: Tiền xử lý dữ liệu

Kỹ năng: Hiểu

Mức độ: Khó

Phần nội dung câu hỏi:

Phát hiện và xử lý các mâu thuẫn đối với giá trị dữ liệu thường được thực hiện thông qua quá trình nào?

Các đáp án:

| | |
|---|--------------------------------------|
| A | Hồi quy (Regression) |
| B | Phân khoảng (Binning) |
| C | Phân tích các cụm (Cluster Analysis) |
| D | Tích hợp dữ liệu |

Đáp án đúng: D

Câu hỏi số: 96

Chương: Tiền xử lý dữ liệu

Kỹ năng: Đánh giá

Mức độ: Dễ

Phần nội dung câu hỏi:

Cần xác định các thực thể (identities) trên thực tế từ nhiều nguồn dữ liệu, trong quá trình tích hợp dữ liệu nhằm mục đích gì?

Các đáp án:

| | |
|---|-------------------------------------------------------------|
| A | Để phân tích dữ liệu theo các cụm. |
| B | Để xác định và xử lý các mâu thuẫn đối với giá trị dữ liệu. |
| C | Để tránh dư thừa dữ liệu. |
| D | Để phân khoảng dữ liệu. |

Đáp án đúng: C

Câu hỏi số: 97

Chương: Tiền xử lý dữ liệu

Kỹ năng: Áp dụng

Mức độ: Khó

Phản nội dung câu hỏi:

Dư thừa dữ liệu (redundant data) thường xuyên xảy ra khi tích hợp dữ liệu từ nhiều nguồn. Các thuộc tính dư thừa có thể được phát hiện bằng phương pháp phân tích tương quan (Correlation analysis) nào sau đây?

Các đáp án:

| | |
|---|----------------------------|
| A | Pearson |
| B | Cosine |
| C | Chi-square |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 98

Chương: Tiền xử lý dữ liệu

Kỹ năng: Áp dụng

Mức độ: Khó

Phản nội dung câu hỏi:

Các phương pháp biến đổi dữ liệu trong tiền xử lý dữ liệu gồm có?

Các đáp án:

| | |
|---|------------------------------------------------|
| A | Làm trơn (Smoothing) và kết hợp (Aggregation). |
| B | Khái quát hóa (Generalization) |
| C | Chuẩn hóa (Normalization) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 99

Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Dễ

Phản nội dung câu hỏi:

Trong quá trình biến đổi dữ liệu, việc loại bỏ nhiễu/lỗi khỏi dữ liệu còn được gọi là phương pháp gì?

Các đáp án:

| | |
|---|---------------|
| A | Làm trơn |
| B | Kết hợp |
| C | Khái quát hóa |
| D | Chuẩn hóa |

Đáp án đúng: A

Câu hỏi số: 100

Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Dễ

Phân nội dung câu hỏi:

Trong quá trình biến đổi dữ liệu, việc tóm tắt dữ liệu, xây dựng các khối dữ liệu (data cubes) còn được gọi bằng phương pháp gì?

Các đáp án:

| | |
|---|---------------|
| A | Làm tròn |
| B | Kết hợp |
| C | Khái quát hóa |
| D | Chuẩn hóa |

Đáp án đúng: B

Câu hỏi số: 101

Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phân nội dung câu hỏi:

Trong quá trình biến đổi dữ liệu, việc xây dựng các phân cấp khái niệm (concept hierarchies) còn được biết đến là phương pháp gì?

Các đáp án:

| | |
|---|---------------|
| A | Làm tròn |
| B | Kết hợp |
| C | Khái quát hóa |
| D | Chuẩn hóa |

Đáp án đúng: C

Câu hỏi số: 102

Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Khó

Phân nội dung câu hỏi:

Trong quá trình biến đổi dữ liệu, việc đưa các giá trị về một khoảng được chỉ định còn được biết đến là phương pháp nào sau đây?

Các đáp án:

| | |
|---|---------------|
| A | Làm tròn |
| B | Kết hợp |
| C | Khái quát hóa |
| D | Chuẩn hóa |

Đáp án đúng: D

Câu hỏi số: 103

Chương: Tiền xử lý dữ liệu

Kỹ năng: Áp dụng

Mức độ: Khó

Phân nội dung câu hỏi:

Biến đổi dữ liệu bằng phương pháp chuẩn hóa (Normalization) bằng cách đưa các giá trị về một khoảng được chỉ định, có thể được thực hiện bởi chuẩn hóa nào sau đây?

Các đáp án:

| | |
|---|-----------------------------|
| A | Chuẩn hóa min-max |
| B | Chuẩn hóa z-score |
| C | Chuẩn hóa bởi thang chia 10 |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 104

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Khó

Phân nội dung câu hỏi:

Phương trình sau thể hiện phương pháp chuẩn hóa gì trong biến đổi dữ liệu?

$$v^{new} = \frac{v^{old} - \min_i}{\max_i - \min_i} (new_max_i - new_min_i) + new_min_i$$

Các đáp án:

| | |
|---|-----------------------------|
| A | Chuẩn hóa min-max |
| B | Chuẩn hóa z-score |
| C | Chuẩn hóa bởi thang chia 10 |
| D | Cả 3 đáp án trên đều đúng |

Đáp án đúng: A

Câu hỏi số: 105

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Khó

Phân nội dung câu hỏi:

Phương trình sau thể hiện phương pháp chuẩn hóa gì trong biến đổi dữ liệu?

$$v^{new} = \frac{v^{old} - \mu_i}{\sigma_i}$$

Trong đó: μ_i, σ_i : giá trị trung bình và độ lệch chuẩn đối với thuộc tính i

Các đáp án:

| | |
|---|-----------------------------|
| A | Chuẩn hóa min-max |
| B | Chuẩn hóa z-score |
| C | Chuẩn hóa bởi thang chia 10 |
| D | Cả 3 đáp án trên đều đúng |

Đáp án đúng: B

Câu hỏi số: 106

Chương: Tiền xử lý dữ liệu

Kỹ năng: Biết

Mức độ: Khó

Phần nội dung câu hỏi:

Phương trình sau thể hiện phương pháp chuẩn hóa gì, trong biến đổi dữ liệu?

$$v^{new} = \frac{v^{old}}{10^j}$$

Trong đó: j là giá trị số nguyên nhỏ nhất sao cho: $\max(\{v^{new}\}) < 1$

Các đáp án:

| | |
|---|-----------------------------|
| A | Chuẩn hóa min-max |
| B | Chuẩn hóa z-score |
| C | Chuẩn hóa bởi thang chia 10 |
| D | Cả 3 đáp án trên đều đúng |

Đáp án đúng: C

Câu hỏi số: 107

Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phần nội dung câu hỏi:

Chiến lược giảm lượng dữ liệu (Data/Numerosity reduction) trong việc giảm bớt dữ liệu có thể được thực hiện bằng cách nào?

Các đáp án:

| | |
|---|----------------------------|
| A | Kết hợp khối dữ liệu |
| B | Nén dữ liệu |
| C | Hồi quy |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 108

Chương: Tiền xử lý dữ liệu

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phân nội dung câu hỏi:

Chiến lược giảm số chiều (Dimensionality reduction) bằng cách loại bỏ bớt các thuộc tính không quan trọng, nhằm đạt được mục đích gì?

Các đáp án:

| | |
|---|------------------------------------------------------------------------------|
| A | Tránh (giảm bớt) ảnh hưởng tiêu cực của số chiều lớn. |
| B | Giúp loại bỏ các thuộc tính không liên quan, và giảm nhiễu/lỗi. |
| C | Giúp giảm chi phí về thời gian và bộ nhớ cần cho quá trình khai phá dữ liệu. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 109

Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Khó

Phân nội dung câu hỏi:

Các kỹ thuật giảm số chiều dữ liệu gồm có?

Các đáp án:

| | |
|---|----------------------------------------------------------------|
| A | Phân tích thành phần chính và lựa chọn tập con các thuộc tính. |
| B | Chuẩn hóa |
| C | Làm tròn |
| D | Khái quát hóa |

Đáp án đúng: A

Câu hỏi số: 110

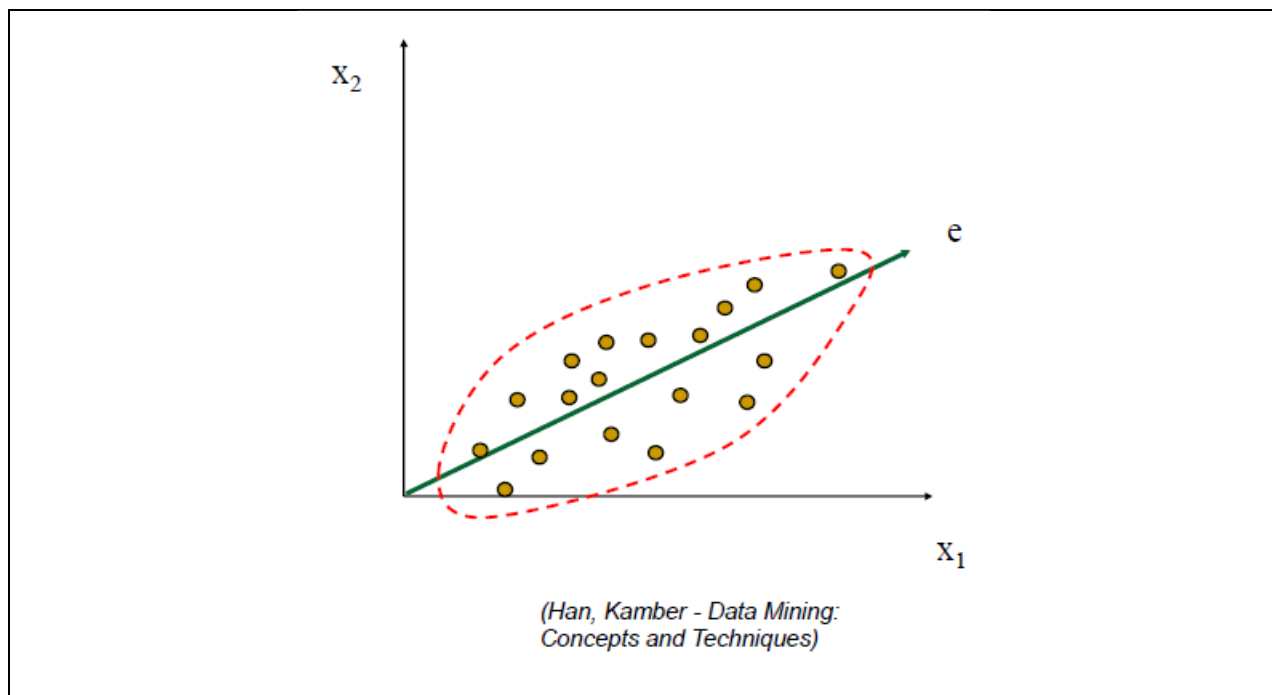
Chương: Tiền xử lý dữ liệu

Kỹ năng: Tổng hợp

Mức độ: Khó

Phân nội dung câu hỏi:

Đồ thị sau thể hiện kỹ thuật nào nhằm giảm số chiều của dữ liệu?



Các đáp án:

| | |
|---|----------------------------------------------------------------|
| A | Phân tích thành phần chính và lựa chọn tập con các thuộc tính. |
| B | Chuẩn hóa |
| C | Làm tròn |
| D | Khái quát hóa |

Đáp án đúng: A

Câu hỏi số: 111

Chương: Tiền xử lý dữ liệu

Kỹ năng: Áp dụng

Mức độ: Khó

Phần nội dung câu hỏi:

Các phương pháp thường được áp dụng cho việc lựa chọn tập con các thuộc tính (Feature subset selection) của dữ liệu gồm có?

Các đáp án:

| | |
|---|-----------------------------------|
| A | Lựa chọn các thuộc tính riêng rẽ. |
| B | Lựa chọn thuộc tính từng bước. |
| C | Loại bỏ thuộc tính từng bước. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 112

Chương: Tiền xử lý dữ liệu

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phần nội dung câu hỏi:

Các phương pháp lấy mẫu dữ liệu gồm có?

Các đáp án:

| | |
|---|-----------------------------------|
| A | Lấy mẫu ngẫu nhiên. |
| B | Lấy mẫu có (hoặc không) thay thế. |
| C | Lấy mẫu phân tầng. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 113

Chương: Phát hiện các luật kết hợp

Kỹ năng: Biết

Mức độ: Trung bình

Phân nội dung câu hỏi:

Phát biểu nào sau đây về bài toán phát hiện luật kết hợp (Association rule mining) là đúng?

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A | Với một tập các giao dịch (transactions) cho trước, cần tìm các luật dự đoán khả năng xuất hiện trong một giao dịch của các mục (items) này dựa trên việc xuất hiện của các mục khác. |
| B | Với một tập các mục (items) cho trước, cần tìm các luật dự đoán khả năng xuất hiện một giao dịch, dựa trên việc xuất hiện của các giao dịch khác. |
| C | Với một tập các thuộc tính cho trước, cần tìm các luật dự đoán khả năng xuất hiện trong một giao dịch của các mục (items) dựa trên việc xuất hiện của các thuộc tính liên quan. |
| D | Với một tập các mối quan hệ ràng buộc cho trước, cần tìm các luật dự đoán khả năng xuất hiện các giao dịch, dựa trên sự xuất hiện của các mục. |

Đáp án đúng: A

Câu hỏi số: 114

Chương: Phát hiện các luật kết hợp

Kỹ năng: Áp dụng

Mức độ: Dễ

Phân nội dung câu hỏi:

Với một tập các giao dịch (transactions) cho trước như bảng sau:

| <i>TID</i> | <i>Items</i> |
|------------|---------------------------|
| 1 | Bread, Milk |
| 2 | Bread, Diaper, Beer, Eggs |
| 3 | Milk, Diaper, Beer, Coke |
| 4 | Bread, Milk, Diaper, Beer |
| 5 | Bread, Milk, Diaper, Coke |

Đâu là ví dụ của luật kết hợp?

Các đáp án:

| | |
|---|-------------------------------|
| A | {Diaper} -> {Beer} |
| B | {Milk, Bread} -> {Eggs, Coke} |
| C | {Beer, Bread} -> {Milk} |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: A

Câu hỏi số: 115

Chương: Phát hiện các luật kết hợp

Kỹ năng: Biết

Mức độ: Dễ

Phần nội dung câu hỏi:

Một tập hợp gồm một hoặc nhiều mục như {Milk, Bread, Diaper},... còn được gọi là gì?

Các đáp án:

| | |
|---|----------------------|
| A | Tổng số hỗ trợ |
| B | Độ hỗ trợ |
| C | Tập mục |
| D | Tập mục thường xuyên |

Đáp án đúng: C

Câu hỏi số: 116

Chương: Phát hiện các luật kết hợp

Kỹ năng: Biết

Mức độ: Dễ

Phần nội dung câu hỏi:

Số lần xuất hiện của một tập mục còn được gọi là gì?

Các đáp án:

| | |
|---|----------------------|
| A | Tổng số hỗ trợ |
| B | Độ hỗ trợ |
| C | Tập mục |
| D | Tập mục thường xuyên |

Đáp án đúng: A

Câu hỏi số: 117

Chương: Phát hiện các luật kết hợp

Kỹ năng: Biết

Mức độ: Khó

Phần nội dung câu hỏi:

Tỷ lệ các giao dịch chứa một tập mục còn được gọi là gì?

Các đáp án:

| | |
|---|----------------------|
| A | Tổng số hỗ trợ |
| B | Độ hỗ trợ |
| C | Tập mục |
| D | Tập mục thường xuyên |

Đáp án đúng: C

Câu hỏi số: 118

Chương: Phát hiện các luật kết hợp

Kỹ năng: Hiểu

Mức độ: Trung bình

Phân nội dung câu hỏi:

Một tập mục mà độ hỗ trợ lớn hơn hoặc bằng một giá trị ngưỡng minsup còn được gọi là gì?

Các đáp án:

| | |
|---|----------------------|
| A | Tổng số hỗ trợ |
| B | Độ hỗ trợ |
| C | Tập mục |
| D | Tập mục thường xuyên |

Đáp án đúng: D

Câu hỏi số: 119

Chương: Phát hiện các luật kết hợp

Kỹ năng: Hiểu

Mức độ: Trung bình

Phân nội dung câu hỏi:

Một biểu thức kéo theo có dạng: $X \rightarrow Y$, trong đó X và Y là các tập mục, còn được gọi là gì?

Các đáp án:

| | |
|---|----------------|
| A | Tổng số hỗ trợ |
| B | Độ hỗ trợ |
| C | Tập mục |
| D | Luật kết hợp |

Đáp án đúng: D

Câu hỏi số: 120

Chương: Phát hiện các luật kết hợp

Kỹ năng: Hiểu

Mức độ: Trung bình

Phân nội dung câu hỏi:

Với một tập các giao dịch T, mục đích của bài toán phát hiện luật kết hợp là tìm ra tất cả các luật thỏa mãn yêu cầu nào sau đây?

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------------|
| A | Độ hỗ trợ \geq giá trị ngưỡng minsup VÀ độ tin cậy \geq giá trị ngưỡng minconf. |
| B | Độ hỗ trợ \geq giá trị ngưỡng minsup HOẶC độ tin cậy \geq giá trị ngưỡng minconf. |
| C | Độ tin cậy \geq giá trị ngưỡng minsup VÀ độ hỗ trợ \geq giá trị ngưỡng minconf. |
| D | Độ tin cậy \geq giá trị ngưỡng minsup HOẶC độ hỗ trợ \geq giá trị ngưỡng minconf. |

Đáp án đúng: A

Câu hỏi số: 121

Chương: Phát hiện các luật kết hợp

Kỹ năng: Hiểu

Mức độ: Khó

Phân nội dung câu hỏi:

Cách tiếp cận vét cạn (Brute-force) trong bài toán phát hiện luật kết hợp có đặc điểm gì?

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------------|
| A | Liệt kê tất cả các luật kết hợp có thể. |
| B | Tính toán độ hỗ trợ và độ tin cậy cho mỗi luật. |
| C | Loại bỏ đi các luật có độ hỗ trợ nhỏ hơn minsup hoặc có độ tin cậy nhỏ hơn minconf. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 122

Chương: Phát hiện các luật kết hợp

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phân nội dung câu hỏi:

Nhận xét nào là đúng, với bài toán phát hiện luật kết hợp như sau:

| TID | Items |
|-----|---------------------------|
| 1 | Bread, Milk |
| 2 | Bread, Diaper, Beer, Eggs |
| 3 | Milk, Diaper, Beer, Coke |
| 4 | Bread, Milk, Diaper, Beer |
| 5 | Bread, Milk, Diaper, Coke |

Các luật kết hợp:

$\{Milk, Diaper\} \rightarrow \{Beer\}$ ($s=0.4, c=0.67$)
 $\{Milk, Beer\} \rightarrow \{Diaper\}$ ($s=0.4, c=1.0$)
 $\{Diaper, Beer\} \rightarrow \{Milk\}$ ($s=0.4, c=0.67$)
 $\{Beer\} \rightarrow \{Milk, Diaper\}$ ($s=0.4, c=0.67$)
 $\{Diaper\} \rightarrow \{Milk, Beer\}$ ($s=0.4, c=0.5$)
 $\{Milk\} \rightarrow \{Diaper, Beer\}$ ($s=0.4, c=0.5$)

Các đáp án:

| | |
|---|-----------------------------------------------------------------------------------|
| A | Tất cả các luật trên đều là sự phân tách thành 2 tập con của cùng 1 tập mục. |
| B | Các luật sinh ra từ cùng một tập mục sẽ có cùng độ hỗ trợ nhưng có thể khác về độ |

| | |
|---|-------------------------------------------------------------------------------------------------|
| | tin cậy. |
| C | Trong quá trình phát hiện luật kết hợp, có thể tách riêng 2 yêu cầu về độ hỗ trợ và độ tin cậy. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 123

Chương: Phát hiện các luật kết hợp

Kỹ năng: Áp dụng

Mức độ: Dễ

Phân nội dung câu hỏi:

Quá trình phát hiện luật kết hợp sẽ gồm bao nhiêu bước quan trọng?

Các đáp án:

| | |
|---|---|
| A | 5 |
| B | 4 |
| C | 3 |
| D | 2 |

Đáp án đúng: D

Câu hỏi số: 124

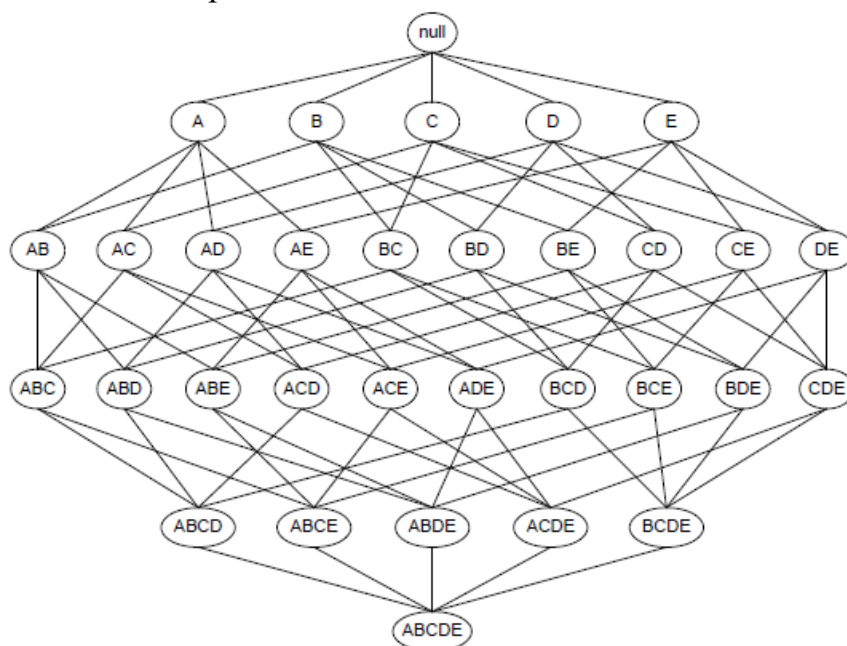
Chương: Phát hiện các luật kết hợp

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phân nội dung câu hỏi:

Theo Lattice biểu diễn các tập mục cần xét:



Với d mục, thì phải xét đến bao nhiêu các tập mục có thể?

Các đáp án:

| | |
|---|-----------|
| A | 2^d |
| B | $2^d - 1$ |
| C | $2^d - 1$ |
| D | d^2 |

Đáp án đúng: A

Câu hỏi số: 125

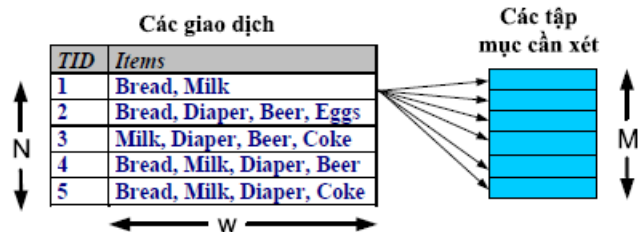
Chương: Phát hiện các luật kết hợp

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

Việc sinh ra các tập mục cần xét theo phương pháp vét cạn (Brute-force) phải thỏa mãn điều kiện gì?



Các đáp án:

| | |
|---|---------------------------------------------------------------------------|
| A | Mỗi tập mục trong Lattice đều được xét. |
| B | Tính độ hỗ trợ của mỗi tập mục, bằng cách duyệt qua tất cả các giao dịch. |
| C | Với mỗi giao dịch, so sánh nó với mỗi tập mục được xét. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 126

Chương: Phát hiện các luật kết hợp

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

Có các chiến lược sinh tập mục thường xuyên nào?

Các đáp án:

| | |
|---|------------------------------------------------------------------------|
| A | Giảm bớt số lượng các tập mục thường xuyên (M). |
| B | Giảm bớt số lượng các giao dịch cần xét (N). |
| C | Giảm bớt số lượng các so sánh giữa các tập mục và các giao dịch (N.M). |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 127

Chương: Phát hiện các luật kết hợp

Kỹ năng: Áp dụng

Mức độ: Khó

Phân nội dung câu hỏi:

Trong chiến lược giảm bớt số lượng các tập mục cần xét (N), nguyên tắc của giải thuật nào dựa trên độ hỗ trợ?

Các đáp án:

| | |
|---|----------------------|
| A | Apriori |
| B | FP-Growth |
| C | FP-Tree |
| D | FP-Tree có điều kiện |

Đáp án đúng: A

Câu hỏi số: 128

Chương: Phát hiện các luật kết hợp

Kỹ năng: Tổng hợp

Mức độ: Dễ

Phân nội dung câu hỏi:

Phải biểu nào sau đây là đúng với giải thuật Apriori – Loại bỏ (prunning)?

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------------------------------------|
| A | Nếu một tập mục là thường xuyên, thì tất cả các tập con của nó đều là tập mục thường xuyên. |
| B | Nếu một tập mục là không thường xuyên, thì tất cả các tập cha của nó đều là các tập mục không thường xuyên. |
| C | Độ hỗ trợ của một tập mục nhỏ hơn độ hỗ trợ của các tập con của nó. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 129

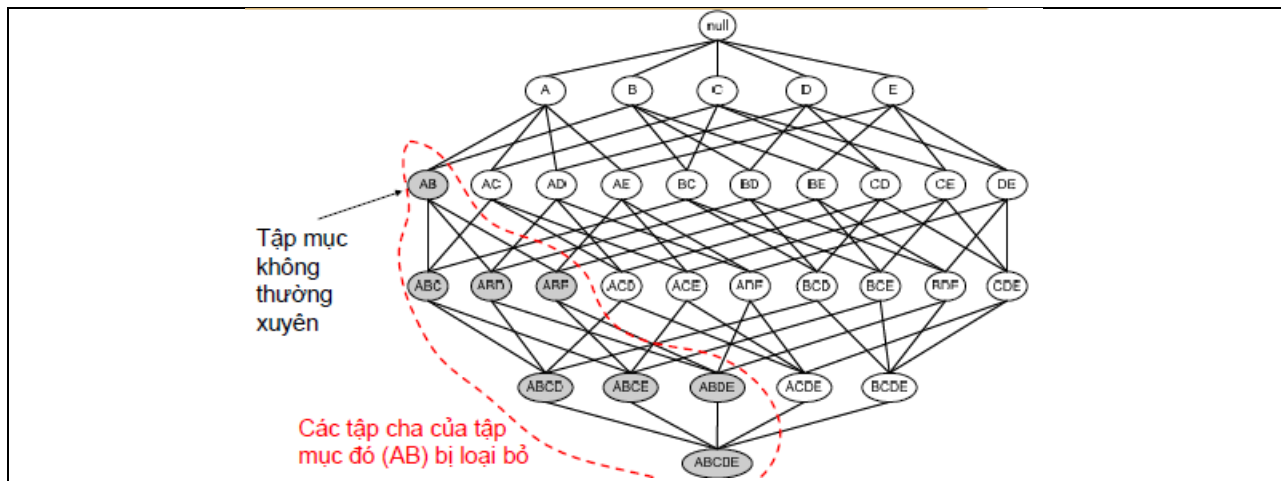
Chương: Phát hiện các luật kết hợp

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phân nội dung câu hỏi:

Đồ hình sau thể hiện giải thuật nào?



Các đáp án:

| | |
|---|-------------------------------------|
| A | Apriori: Loại bỏ dựa trên độ hỗ trợ |
| B | FP-Growth |
| C | FP-Tree |
| D | FP-Tree có điều kiện |

Đáp án đúng: A

Câu hỏi số: 130

Chương: Phát hiện các luật kết hợp

Kỹ năng: Biết

Mức độ: Khó

Phân nội dung câu hỏi:

Trong giải thuật Apriori, các tập mục thường xuyên chỉ chứa 1 mục và k được gán bằng 1 nằm ở mức bao nhiêu?

Các đáp án:

| | |
|---|---|
| A | 2 |
| B | 1 |
| C | 3 |
| D | 4 |

Đáp án đúng: B

Câu hỏi số: 131

Chương: Phát hiện các luật kết hợp

Kỹ năng: Phân tích

Mức độ: Khó

Phân nội dung câu hỏi:

Giải thuật Apriori sẽ dừng lại khi nào?

Các đáp án:

| | |
|---|--------------------------------------------------------|
| A | Khi không có thêm bất kỳ tập mục thường xuyên nào mới. |
|---|--------------------------------------------------------|

| | |
|---|--------------------------------------------------------------|
| B | Khi các tập mục thường xuyên đạt đến một mức cho trước. |
| C | Khi đạt được độ hỗ trợ cần thiết. |
| D | Khi số lượng các so sánh được giảm bớt đến mức độ cần thiết. |

Đáp án đúng: A

Câu hỏi số: 132

Chương: Phát hiện các luật kết hợp

Kỹ năng: Phân tích

Mức độ: Khó

Phần nội dung câu hỏi:

Để giảm bớt số lượng các so sánh, và lưu các tập mục cần xét, giải pháp nào được sử dụng?

Các đáp án:

| | |
|---|-------------------------------------|
| A | Apriori: Loại bỏ dựa trên độ hỗ trợ |
| B | FP-Growth |
| C | Cấu trúc hàm băm (Hash structure) |
| D | FP-Tree có điều kiện |

Đáp án đúng: C

Câu hỏi số: 133

Chương: Phát hiện các luật kết hợp

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phần nội dung câu hỏi:

Các yếu tố nào ảnh hưởng đến độ phức tạp trong giải pháp Apriori?

Các đáp án:

| | |
|---|---------------------------------|
| A | Minsup: lựa chọn ngưỡng giá trị |
| B | Số lượng các mục trong CSDL |
| C | Kích thước của CSDL |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 134

Chương: Phát hiện các luật kết hợp

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong giải pháp Apriori, điều gì có thể làm tăng số lượng các tập mục phải xét và độ dài (kích thước) tối đa của các tập mục thường xuyên?

Các đáp án:

| | |
|---|-----------------------------------------------------|
| A | Minsup: lựa chọn ngưỡng giá trị quá thấp. |
| B | Số lượng các mục trong CSDL cố định. |
| C | Kích thước của CSDL không đổi. |
| D | Kích thước trung bình của các giao dịch giảm xuống. |

Đáp án đúng: A

Câu hỏi số: 135

Chương: Phát hiện các luật kết hợp

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong giải pháp Apriori, khi nào chúng ta cần thêm bộ nhớ để lưu giá trị độ hỗ trợ đối với mỗi mục?

Các đáp án:

| | |
|---|-----------------------------------------------------|
| A | Minsup: lựa chọn ngưỡng giá trị quá thấp. |
| B | Số lượng các mục trong các giao dịch tăng lên. |
| C | Kích thước của CSDL không đổi. |
| D | Kích thước trung bình của các giao dịch giảm xuống. |

Đáp án đúng: B

Câu hỏi số: 136

Chương: Phát hiện các luật kết hợp

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phần nội dung câu hỏi:

Giải thuật Apriori duyệt CSDL nhiều lần, làm tăng chi phí tính toán của Apriori khi số lượng các giao dịch tăng lên. Điều này phụ thuộc yếu tố nào sau đây?

Các đáp án:

| | |
|---|-----------------------------------------------------|
| A | Minsup: lựa chọn ngưỡng giá trị quá thấp. |
| B | Số lượng các mục trong CSDL cố định. |
| C | Kích thước của CSDL không đổi. |
| D | Kích thước trung bình của các giao dịch giảm xuống. |

Đáp án đúng: C

Câu hỏi số: 137

Chương: Phát hiện các luật kết hợp

Kỹ năng: Tổng hợp

Mức độ: Khó

Phần nội dung câu hỏi:

Trong giải thuật Apriori, khi độ dài tối đa của các tập mục thường xuyên tăng lên, chi phí

duyet hàm băm cũng sẽ tăng lên. Điều này phụ thuộc yếu tố nào sau đây?

Các đáp án:

| | |
|---|------------------------------------------|
| A | Lựa chọn giá trị ngưỡng minsup. |
| B | Số lượng các mục trong CSDL. |
| C | Kích thước của CSDL. |
| D | Kích thước trung bình của các giao dịch. |

Đáp án đúng: D

Câu hỏi số: 138

Chương: Phát hiện các luật kết hợp

Kỹ năng: Tổng hợp

Mức độ: Dễ

Phần nội dung câu hỏi:

Có bao nhiêu cách biểu diễn các tập mục thường xuyên?

Các đáp án:

| | |
|---|---|
| A | 1 |
| B | 2 |
| C | 3 |
| D | 4 |

Đáp án đúng: B

Câu hỏi số: 139

Chương: Phát hiện các luật kết hợp

Kỹ năng: Tổng hợp

Mức độ: Khó

Phần nội dung câu hỏi:

Phát biểu nào sau đây về tập mục thường xuyên lớn nhất là đúng?

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------------------------------|
| A | Một tập mục thường xuyên lớn nhất, nếu mọi tập cha (superset) của nó đều là tập mục không thường xuyên. |
| B | Một tập mục thường xuyên lớn nhất, nếu mọi tập con của nó đều là tập mục không thường xuyên. |
| C | Một tập mục thường xuyên lớn nhất, nếu mọi tập cha (superset) của nó đều là tập mục thường xuyên. |
| D | Một tập mục thường xuyên lớn nhất, nếu mọi tập con của nó đều là tập mục thường xuyên. |

Đáp án đúng: A

Câu hỏi số: 140

Chương: Phát hiện các luật kết hợp

Kỹ năng: Tổng hợp

Mức độ: Khó

Phản nội dung câu hỏi:

Phát biểu nào sau đây là đúng về các tập mục thường xuyên đóng?

Các đáp án:

| | |
|---|-----------------------------------------------------------------------------------------------|
| A | Một tập mục thường xuyên là đóng, nếu không có tập con nào của nó có cùng độ hỗ trợ với nó. |
| B | Một tập mục thường xuyên là đóng, nếu không có tập cha nào của nó có cùng độ hỗ trợ với nó. |
| C | Một tập mục thường xuyên là đóng, nếu không có tập cha nào của nó có cùng độ phức tạp với nó. |
| D | Một tập mục thường xuyên là đóng, nếu không có tập con nào của nó có cùng độ phức tạp với nó. |

Đáp án đúng: B

Câu hỏi số: 141

Chương: Phát hiện các luật kết hợp

Kỹ năng: Biết

Mức độ: Trung bình

Phản nội dung câu hỏi:

Phát biểu nào sau đây về giải thuật FP-Growth là đúng?

Các đáp án:

| | |
|---|-----------------------------------------------------------------------------------------|
| A | FP-Growth là một phương pháp xác định các tập mục thường xuyên. |
| B | FP-Growth biểu diễn dữ liệu của các giao dịch bằng một cấu trúc dữ liệu gọi là FP-Tree. |
| C | FP-Growth sử dụng cấu trúc FP-Tree để xác định trực tiếp các tập mục thường xuyên. |
| D | Cả 3 đáp án trên đều đúng |

Đáp án đúng: D

Câu hỏi số: 142

Chương: Phát hiện các luật kết hợp

Kỹ năng: Hiểu

Mức độ: Khó

Phản nội dung câu hỏi:

Trong giải thuật FP-Growth, mỗi giao dịch là một thành phần nào của cây FP-Tree?

Các đáp án:

| | |
|---|----------------------------|
| A | Đường đi (path) trong cây. |
| B | Nhánh (branch) của cây. |
| C | Gốc của cây. |

| | |
|---|---------------------|
| D | Kích thước của cây. |
|---|---------------------|

Đáp án đúng: A

Câu hỏi số: 143

Chương: Phát hiện các luật kết hợp

Kỹ năng: Tổng hợp

Mức độ: Khó

Phần nội dung câu hỏi:

| |
|-------------------------------------------|
| Phát biểu nào sau đây về FP-Tree là đúng? |
|-------------------------------------------|

Các đáp án:

| | |
|---|--------------------------------------------------------------------------------------------|
| A | Hai giao dịch có chứa cùng một số các mục, thì đường đi của chúng sẽ có phần (đoạn) chung. |
| B | Càng nhiều các đường đi có phần chung, thì việc biểu diễn bằng FP-Tree sẽ càng phức tạp. |
| C | FP-Tree xây dựng nhiều đường đi (path) trong cây với mỗi giao dịch. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: A

Câu hỏi số: 144

Chương: Phát hiện các luật kết hợp

Kỹ năng: Tổng hợp

Mức độ: Khó

Phần nội dung câu hỏi:

| |
|---------------------------------------------------------------------------|
| Có bao nhiêu nút gốc (được biểu diễn bởi ký hiệu null) trong một FP-Tree? |
|---------------------------------------------------------------------------|

Các đáp án:

| | |
|---|---|
| A | 4 |
| B | 3 |
| C | 2 |
| D | 1 |

Đáp án đúng: D

Câu hỏi số: 145

Chương: Phát hiện các luật kết hợp

Kỹ năng: Biết

Mức độ: Khó

Phần nội dung câu hỏi:

| |
|-------------------------------------------------------------------------------------------------------------------|
| Trong giải thuật FP-Growth, cơ sở dữ liệu các giao dịch được duyệt lần thứ nhất, để xác định điều gì của mỗi mục? |
|-------------------------------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|-------------|
| A | Độ phức tạp |
|---|-------------|

| | |
|---|----------------------|
| B | Độ hỗ trợ |
| C | Các mục thường xuyên |
| D | Xây dựng FP-Tree |

Đáp án đúng: B

Câu hỏi số: 146

Chương: Phát hiện các luật kết hợp

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong giải thuật FP-Growth, cơ sở dữ liệu các giao dịch được duyệt lần thứ hai, để xác định điều gì của mỗi mục?

Các đáp án:

| | |
|---|----------------------|
| A | Độ phức tạp |
| B | Độ hỗ trợ |
| C | Các mục thường xuyên |
| D | Xây dựng FP-Tree |

Đáp án đúng: D

Câu hỏi số: 147

Chương: Phát hiện các luật kết hợp

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong giải thuật FP-Growth, các mục thường xuyên (frequent items) được sắp xếp như thế nào?

Các đáp án:

| | |
|---|--------------------------------------|
| A | Theo thứ tự tăng dần về độ hỗ trợ. |
| B | Theo thứ tự tăng dần về độ phức tạp. |
| C | Theo thứ tự giảm dần về độ hỗ trợ. |
| D | Theo thứ tự giảm dần về độ phức tạp. |

Đáp án đúng: C

Câu hỏi số: 148

Chương: Phát hiện các luật kết hợp

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong giải thuật FP-Growth, mỗi giao dịch được biểu diễn bằng một đường đi trong FP-Tree, nên chúng ta có thể xác định các tập mục thường xuyên kết thúc bởi một mục bằng

cách nào?

Các đáp án:

| | |
|---|-----------------------------------------------------------------------|
| A | Tìm các tập mục thường xuyên trực tiếp từ FP-Tree. |
| B | Tìm các tập mục không thường xuyên từ mức lá đến mức gốc (bottom-up). |
| C | Duyệt các đường đi chứa mục đó. |
| D | Tìm các tập mục thường xuyên chứa mục đó. |

Đáp án đúng: C

Câu hỏi số: 149

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Biết

Mức độ: Dễ

Phần nội dung câu hỏi:

Phát biểu nào sau đây về bài toán phân lớp (Classification) là đúng?

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------------------------------------|
| A | Một tập các ví dụ/bản ghi được gọi là tập huấn luyện/học. |
| B | Mỗi tập bản ghi biểu diễn bằng một tập các thuộc tính, trong đó có một thuộc tính phân lớp (class attribute). |
| C | Tìm/học một hàm cho thuộc tính phân lớp (hàm phân lớp) đối với các giá trị của các thuộc tính khác. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 150

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Biết

Mức độ: Trung bình

Phần nội dung câu hỏi:

Tập kiểm thử (test set) trong bài toán phân lớp được hình thành như thế nào?

Các đáp án:

| | |
|---|------------------------------------------------------------------------------------------------------|
| A | Sử dụng một tập các ví dụ khác với các ví dụ học để kiểm tra độ chính xác của hàm phân lớp học được. |
| B | Tập dữ liệu ban đầu được chia thành các tập kiểm thử hàm phân lớp học được gọi là test set. |
| C | Sử dụng các ví dụ giống với các tập học để kiểm tra độ chính xác của hàm phân lớp. |
| D | Tập dữ liệu ban đầu được chia thành 2 tập, trong đó có tập test set để học hàm phân lớp. |

Đáp án đúng: A

Câu hỏi số: 151

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phần nội dung câu hỏi:

Học một hàm mục tiêu có giá trị rời rạc trong một miền giá trị với một tập các nhãn lớp xác định trước, thuộc bài toán nào?

Các đáp án:

| | |
|---|--------------------|
| A | Bài toán dự đoán. |
| B | Bài toán hồi quy. |
| C | Bài toán phân lớp. |
| D | Bài toán đệ quy. |

Đáp án đúng: C

Câu hỏi số: 152

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phần nội dung câu hỏi:

Học một hàm mục tiêu có giá trị liên tục trong miền giá trị gồm các giá trị số thực, thuộc bài toán nào?

Các đáp án:

| | |
|---|----------------------------|
| A | Bài toán dự đoán/hồi quy. |
| B | Bài toán đệ quy. |
| C | Bài toán phân lớp. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: A

Câu hỏi số: 153

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

Với mỗi ví dụ cần phân loại trong bài toán phân lớp, hệ thống cần phải làm gì?

Các đáp án:

| | |
|---|----------------------------------------------------|
| A | Hệ thống học một hàm mục tiêu có giá trị rời rạc. |
| B | Hệ thống học một hàm mục tiêu có giá trị liên tục. |
| C | Hệ thống xác định nhãn lớp của nó. |
| D | Hệ thống xác định giá trị dự đoán của nó. |

Đáp án đúng: C

Câu hỏi số: 154

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

Phương pháp học nào trong đó mỗi ví dụ học gồm 2 phần: mô tả (biểu diễn) của ví dụ học, và nhãn lớp (hoặc giá trị đầu ra mong muốn) của ví dụ học đó?

Các đáp án:

| | |
|---|------------------------|
| A | Học không có giám sát. |
| B | Học có giám sát. |
| C | Học phân lớp |
| D | Học phân cụm |

Đáp án đúng: B

Câu hỏi số: 155

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

Phương pháp học nào trong đó mỗi ví dụ học chỉ chứa mô tả (biểu diễn) của ví dụ học đó – mà không có bất kỳ thông tin nào về nhãn lớp hay giá trị đầu ra mong muốn của ví dụ đó?

Các đáp án:

| | |
|---|------------------------|
| A | Học không có giám sát. |
| B | Học có giám sát. |
| C | Học phân lớp |
| D | Học dự đoán/hồi quy |

Đáp án đúng: A

Câu hỏi số: 156

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

Công thức nào sau đây biểu diễn bài toán học phân lớp?

Các đáp án:

| | |
|---|------------------------------------------------------------------------------------------------------------------------|
| A | $D_{\text{train}} = \{(\langle \text{Biểu_diễn_của_}x \rangle, \langle \text{Nhãn_lớp_của_}x \rangle)\}$ |
| B | $D_{\text{train}} = \{(\langle \text{Biểu_diễn_của_}x \rangle, \langle \text{Giá_trị_đầu_ra_của_}x \rangle)\}$ |
| C | Tập học $D_{\text{train}} = \{(\langle \text{Biểu_diễn_của_}x \rangle)\}$ |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: A

Câu hỏi số: 157

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Hiểu

Mức độ: Trung bình

Phản nội dung câu hỏi:

Công thức nào sau đây biểu diễn bài toán học dự đoán/hồi quy?

Các đáp án:

| | |
|---|------------------------------------------------------------------------------------------------------------------------|
| A | $D_{\text{train}} = \{(\langle \text{Biểu_diễn_của_x} \rangle, \langle \text{Nhãn_lớp_của_x} \rangle)\}$ |
| B | $D_{\text{train}} = \{(\langle \text{Biểu_diễn_của_x} \rangle, \langle \text{Giá_trị_đầu_ra_của_x} \rangle)\}$ |
| C | Tập học $D_{\text{train}} = \{(\langle \text{Biểu_diễn_của_x} \rangle)\}$ |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: B

Câu hỏi số: 158

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Hiểu

Mức độ: Trung bình

Phản nội dung câu hỏi:

Công thức nào sau đây biểu diễn bài toán học phân cụm (Clustering problem)?

Các đáp án:

| | |
|---|------------------------------------------------------------------------------------------------------------------------|
| A | $D_{\text{train}} = \{(\langle \text{Biểu_diễn_của_x} \rangle, \langle \text{Nhãn_lớp_của_x} \rangle)\}$ |
| B | $D_{\text{train}} = \{(\langle \text{Biểu_diễn_của_x} \rangle, \langle \text{Giá_trị_đầu_ra_của_x} \rangle)\}$ |
| C | Tập học $D_{\text{train}} = \{(\langle \text{Biểu_diễn_của_x} \rangle)\}$ |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: C

Câu hỏi số: 159

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phản nội dung câu hỏi:

Giả sử chúng ta có một thí nghiệm mà kết quả của nó mang tính ngẫu nhiên. Khi đó, tập hợp tất cả các kết quả có thể xảy ra được gọi là gì?

Các đáp án:

| | |
|---|----------------------------|
| A | Không gian các khả năng S. |
| B | Sự kiện E. |
| C | Không gian các sự kiện W. |
| D | Biến ngẫu nhiên A. |

Đáp án đúng: A

Câu hỏi số: 160

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phân nội dung câu hỏi:

Giả sử chúng ta có một thí nghiệm mà kết quả của nó mang tính ngẫu nhiên. Khi đó, một tập con của không gian các khả năng, được gọi là gì?

Các đáp án:

| | |
|---|----------------------------|
| A | Không gian các khả năng S. |
| B | Sự kiện E. |
| C | Không gian các sự kiện W. |
| D | Biến ngẫu nhiên A. |

Đáp án đúng: B

Câu hỏi số: 161

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phân nội dung câu hỏi:

Giả sử chúng ta có một thí nghiệm mà kết quả của nó mang tính ngẫu nhiên. Khi đó, không gian (thế giới) mà các kết quả của sự kiện có thể xảy ra, được gọi là gì?

Các đáp án:

| | |
|---|----------------------------|
| A | Không gian các khả năng S. |
| B | Sự kiện E. |
| C | Không gian các sự kiện W. |
| D | Biến ngẫu nhiên A. |

Đáp án đúng: C

Câu hỏi số: 162

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phân nội dung câu hỏi:

Giả sử chúng ta có một thí nghiệm mà kết quả của nó mang tính ngẫu nhiên. Khi đó, một biến ngẫu nhiên biểu diễn (diễn đạt) một sự kiện, và có một mức độ về khả năng xảy ra sự kiện này, được gọi là gì?

Các đáp án:

| | |
|---|----------------------------|
| A | Không gian các khả năng S. |
| B | Sự kiện E. |

| | |
|---|---------------------------|
| C | Không gian các sự kiện W. |
| D | Biến ngẫu nhiên A. |

Đáp án đúng: D

Câu hỏi số: 163

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

| |
|--------------------------------------------------------------------------------------------------------------------------------|
| <p>Trong định lý Bayes $P(h D) = \frac{P(D h).P(h)}{P(D)}$, xác suất trước của giả thiết là thành phần nào?</p> |
|--------------------------------------------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|-----------|
| A | P (h) |
| B | P (D) |
| C | P (D h) |
| D | P (h D) |

Đáp án đúng: A

Câu hỏi số: 164

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Khó

Phần nội dung câu hỏi:

| |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Trong định lý Bayes $P(h D) = \frac{P(D h).P(h)}{P(D)}$, xác suất trước của sự kiện tập dữ liệu D được quan sát, là thành phần nào?</p> |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|-----------|
| A | P (h) |
| B | P (D) |
| C | P (D h) |
| D | P (h D) |

Đáp án đúng: B

Câu hỏi số: 165

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Khó

Phần nội dung câu hỏi:

$$P(h | D) = \frac{P(D | h).P(h)}{P(D)}$$

Trong định lý Bayes, xác suất trước của sự kiện tập dữ liệu D được quan sát, nếu biết rằng giả thiết h là đúng, là thành phần nào?

Các đáp án:

| | |
|---|-----------|
| A | P (h) |
| B | P (D) |
| C | P (D h) |
| D | P (h D) |

Đáp án đúng: C

Câu hỏi số: 166

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

$$P(h | D) = \frac{P(D | h).P(h)}{P(D)}$$

Trong định lý Bayes, xác suất (có điều kiện) của giả thiết h là đúng, nếu biết rằng tập dữ liệu D được quan sát, là thành phần nào?

Các đáp án:

| | |
|---|-----------|
| A | P (h) |
| B | P (D) |
| C | P (D h) |
| D | P (h D) |

Đáp án đúng: D

Câu hỏi số: 167

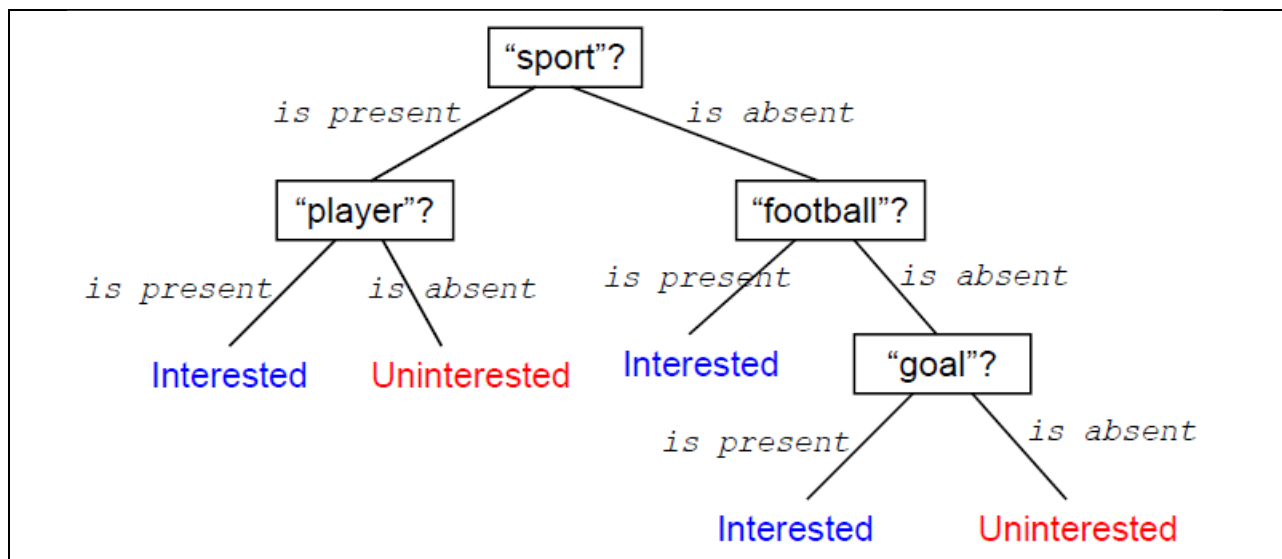
Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong ví dụ sau về học cây quyết định (DT) – những tin tức nào dưới đây chúng ta không quan tâm?



Các đáp án:

| | |
|---|----------------------------------|
| A | (..., "sport",..., "player",...) |
| B | (..., "goal",...) |
| C | (..., "sport",...) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: B

Câu hỏi số: 168

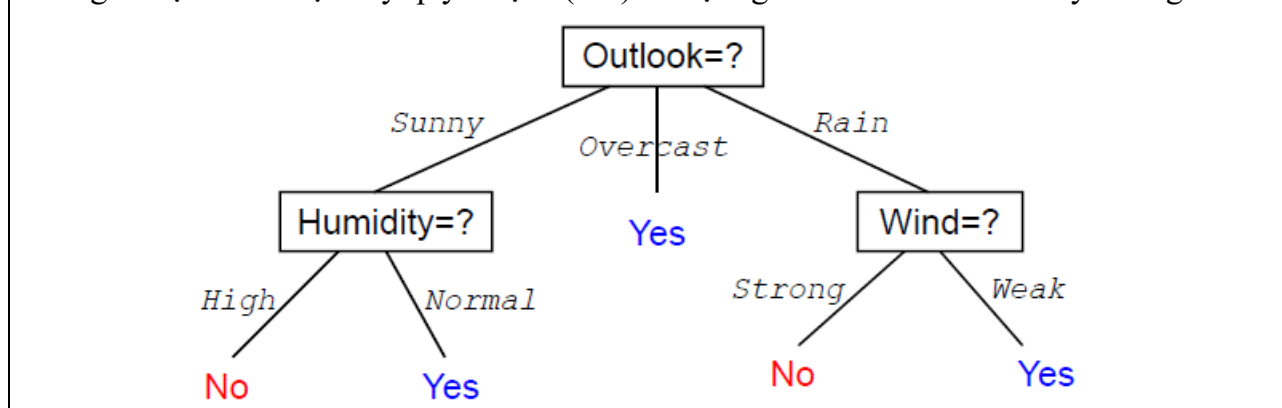
Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Biết

Mức độ: Dễ

Phản nội dung câu hỏi:

Trong ví dụ sau về học cây quyết định (DT) – Một người có chơi Tennis hay không?



Các đáp án:

| | |
|---|---------------------------------------------------------------|
| A | (Outlook=Overcast, Temperature=Hot, Humidity=High, Wind=Weak) |
| B | (Outlook=Rain, Temperature=Mild, Humidity=High, Wind=Strong) |
| C | (Outlook=Sunny, Temperature=Hot, Humidity=High, Wind=Strong) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: A

Câu hỏi số: 169

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Biết

Mức độ: Trung bình

Phân nội dung câu hỏi:

Phát biểu nào sau đây về học cây quyết định (Decision tree – DT – learning) là đúng?

Các đáp án:

| | |
|---|--------------------------------------------------------------------------------------|
| A | Một cây quyết định có thể được biểu diễn bằng một tập các luật IF-THEN. |
| B | Học cây quyết định có thể thực hiện ngay cả với các dữ liệu có chứa nhiễu/lỗi. |
| C | Học cây quyết định là một trong các phương pháp học quy nạp được dùng phổ biến nhất. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 170

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Phân tích

Mức độ: Trung bình

Phân nội dung câu hỏi:

Khi biểu diễn cây quyết định, thành phần nào thể hiện một thuộc tính cần kiểm tra giá trị (an attribute to be tested) đối với các ví dụ?

Các đáp án:

| | |
|---|---------------------------|
| A | Nút trong – Internal node |
| B | Nhánh - Branch |
| C | Nút lá – Leaf node |
| D | Nhãn - Label |

Đáp án đúng: A

Câu hỏi số: 171

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Phân tích

Mức độ: Trung bình

Phân nội dung câu hỏi:

Khi biểu diễn cây quyết định, thành phần nào từ đó sẽ tương ứng với một giá trị có thể của thuộc tính gắn với nút đó?

Các đáp án:

| | |
|---|---------------------------|
| A | Nút trong – Internal node |
| B | Nhánh - Branch |
| C | Nút lá – Leaf node |

| | |
|---|--------------|
| D | Nhãn - Label |
|---|--------------|

Đáp án đúng: B

Câu hỏi số: 172

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phản nội dung câu hỏi:

| |
|-----------------------------------------------------------------------------------------|
| Trong biểu diễn cây quyết định, thành phần nào biểu diễn một phân lớp (classification)? |
|-----------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|---------------------------|
| A | Nút trong – Internal node |
| B | Nhánh - Branch |
| C | Nút lá – Leaf node |
| D | Nhãn - Label |

Đáp án đúng: C

Câu hỏi số: 173

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Phân tích

Mức độ: Trung bình

Phản nội dung câu hỏi:

| |
|-----------------------------------------------------------------------------------------------------|
| Trong giải thuật ID3 – việc xây dựng một cây quyết định theo chiến lược Top-down sẽ bắt đầu từ đâu? |
|-----------------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|---------|
| A | Nút lá |
| B | Nhánh |
| C | Nút gốc |
| D | Cành |

Đáp án đúng: C

Câu hỏi số: 174

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Phân tích

Mức độ: Trung bình

Phản nội dung câu hỏi:

| |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Trong giải thuật ID3 – Thuộc tính kiểm tra (test attribute) là thuộc tính có khả năng phân loại tốt nhất đối với các ví dụ học gắn với thành phần nào sau đây? |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|-------------------|
| A | Nút được kiểm tra |
| B | Nút gốc |

| | |
|---|----------------------------|
| C | Nhánh được kiểm tra |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: A

Câu hỏi số: 175

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Phân tích

Mức độ: Trung bình

Phản nội dung câu hỏi:

Trong giải thuật ID3 – Mỗi thuộc tính được phép xuất hiện tối đa bao nhiêu lần đối với bất kỳ một đường đi nào trong cây?

Các đáp án:

| | |
|---|---|
| A | 4 |
| B | 3 |
| C | 2 |
| D | 1 |

Đáp án đúng: D

Câu hỏi số: 176

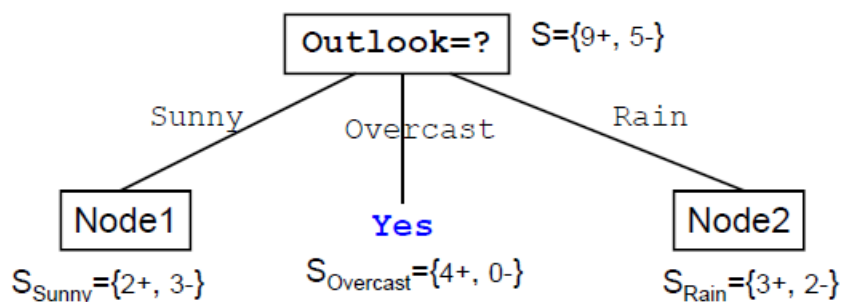
Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Phân tích

Mức độ: Khó

Phản nội dung câu hỏi:

Tại nút gốc, thuộc tính nào trong số $\{Outlook, Temperature, Humidity, Wind\}$ nên được chọn là thuộc tính kiểm tra? (Bằng cách tính giá trị Information Gain của thuộc tính)



Các đáp án:

| | |
|---|-------------------------------|
| A | Gain (S, Outlook) = 0.246 |
| B | Gain (S, Temperature) = 0.029 |
| C | Gain (S, Humidity) = 0.151 |
| D | Gain (S, Wind) = 0.048 |

Đáp án đúng: A

Câu hỏi số: 177

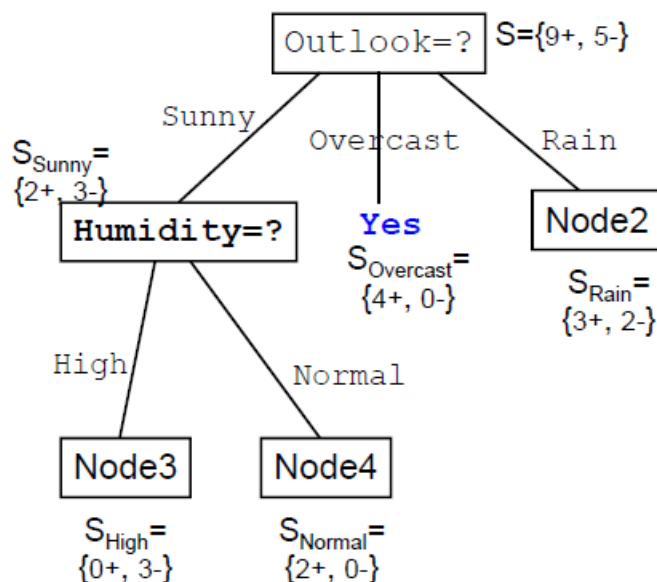
Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Phân tích

Mức độ: Trung bình

Phần nội dung câu hỏi:

Tại nút Node1, thuộc tính nào trong số <Temperature, Humidity, Wind> nên được chọn là thuộc tính kiểm tra?



Các đáp án:

| | |
|---|----------------------------|
| A | Temperature |
| B | Humidity |
| C | Wind |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: B

Câu hỏi số: 178

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Đánh giá

Mức độ: Dễ

Phần nội dung câu hỏi:

Chiến lược tìm kiếm trong học cây quyết định bao gồm những gì?

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------------|
| A | ID3 tìm kiếm trong không gian các giả thiết một cây quyết định phù hợp các ví dụ học. |
| B | ID3 thực hiện chiến lược tìm kiếm từ đơn giản đến phức tạp, bắt đầu với cây rỗng. |
| C | Quá trình tìm kiếm của ID3 được điều khiển bởi độ đo đánh giá Information Gain. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 179

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phân nội dung câu hỏi:

Vấn đề nào sau đây sẽ gặp phải trong học cây quyết định?

Các đáp án:

| | |
|---|--------------------------------------------------------|
| A | Cây quyết định học được quá phù hợp với các ví dụ học. |
| B | Xử lý các thuộc tính có kiểu giá trị liên tục. |
| C | Xử lý các ví dụ học thiếu giá trị thuộc tính. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 180

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phân nội dung câu hỏi:

Tên gọi khác của phương pháp học dựa trên các láng giềng gần nhất (Nearest neighbor learning) là?

Các đáp án:

| | |
|---|----------------------------|
| A | Instance-based learning |
| B | Lazy learning |
| C | Memory-based learning |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 181

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phân nội dung câu hỏi:

Biểu diễn đầu vào của bài toán học dựa trên các láng giềng gần nhất: với mỗi ví dụ x được biểu diễn là một véc-tơ bao nhiêu chiều trong không gian véc-tơ X ?

Các đáp án:

| | |
|---|------------|
| A | n chiều |
| B | N^2 |
| C | $2n$ chiều |
| D | $3n$ chiều |

Đáp án đúng: A

Câu hỏi số: 182

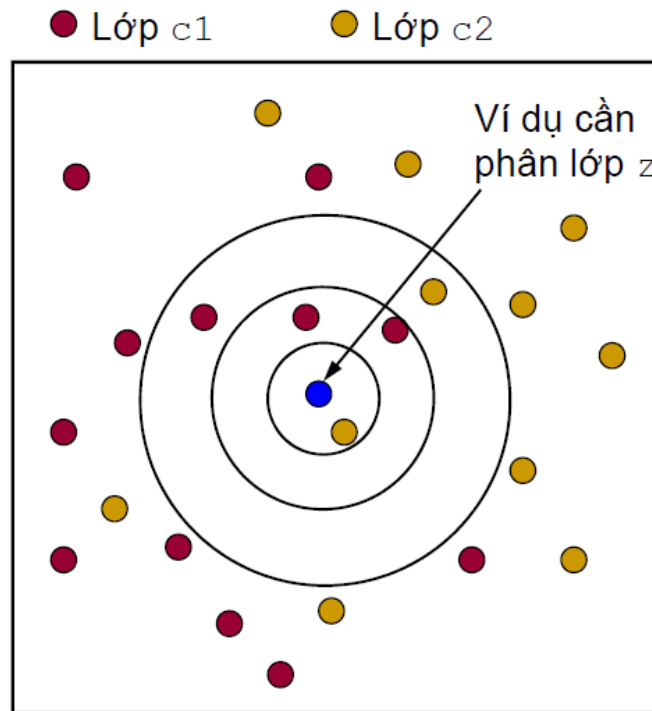
Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phân nội dung câu hỏi:

Ví dụ bài toán phân lớp sau:



Khi nào z được gán vào lớp c1?

Các đáp án:

| | |
|---|-----------------------------|
| A | Xét một láng giềng gần nhất |
| B | Xét ba láng giềng gần nhất |
| C | Xét một láng giềng xa nhất |
| D | Xét ba láng giềng xa nhất |

Đáp án đúng: B

Câu hỏi số: 183

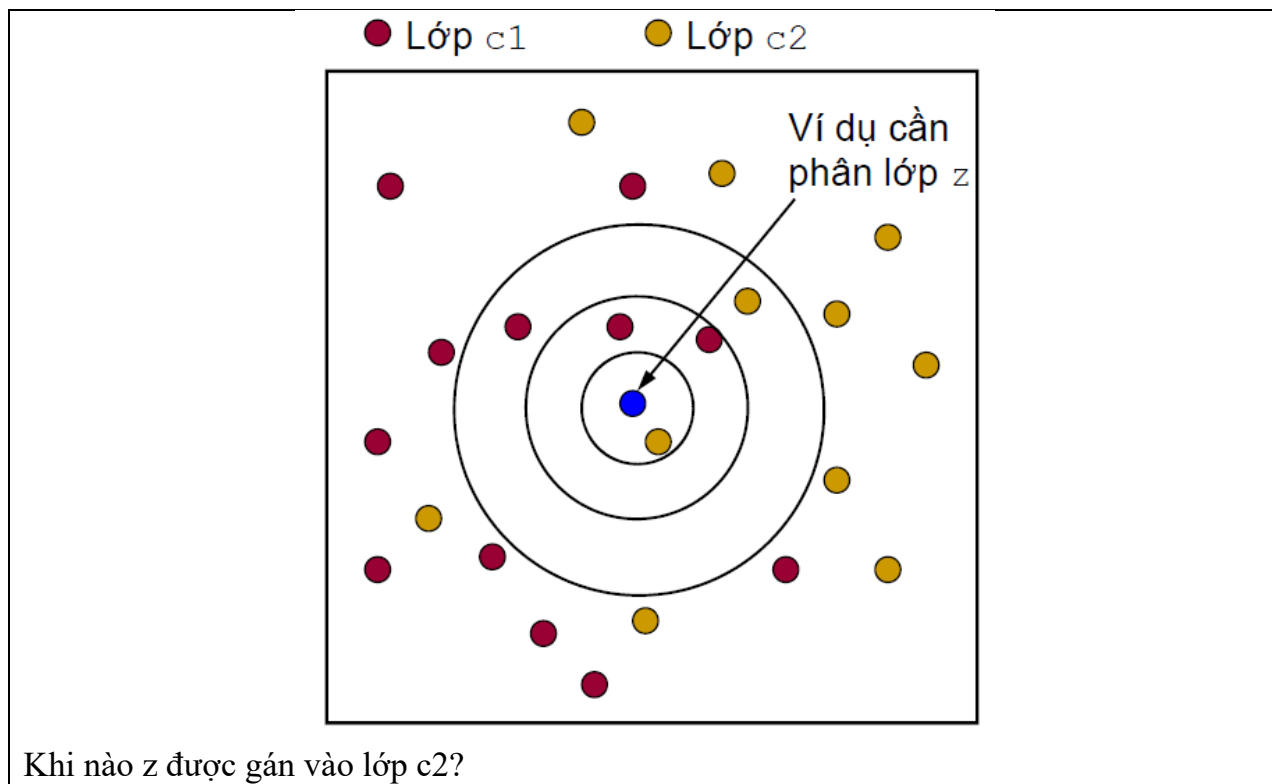
Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phân nội dung câu hỏi:

Ví dụ bài toán phân lớp sau:



Khi nào z được gán vào lớp c2?

Các đáp án:

| | |
|---|-----------------------------|
| A | Xét một láng giềng gần nhất |
| B | Xét ba láng giềng gần nhất |
| C | Xét một láng giềng xa nhất |
| D | Xét ba láng giềng xa nhất |

Đáp án đúng: A

Câu hỏi số: 184

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong giải thuật phân lớp k-NN, mỗi ví dụ học x được biểu diễn bởi mấy thành phần?

Các đáp án:

| | |
|---|---|
| A | 1 |
| B | 2 |
| C | 3 |
| D | 4 |

Đáp án đúng: B

Câu hỏi số: 185

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong giải thuật phân lớp k-NN, quá trình lưu các ví dụ học trong tập học $D = \{x\}$, thuộc giai đoạn nào?

Các đáp án:

| | |
|---|----------------------------|
| A | Giai đoạn học |
| B | Giai đoạn phân lớp |
| C | Giai đoạn dự đoán |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: A

Câu hỏi số: 186

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong giải thuật phân lớp k-NN, quá trình phân lớp cho một ví dụ (mới) z , thuộc giai đoạn nào?

Các đáp án:

| | |
|---|----------------------------|
| A | Giai đoạn học |
| B | Giai đoạn phân lớp |
| C | Giai đoạn dự đoán |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: B

Câu hỏi số: 187

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phần nội dung câu hỏi:

Các bước nào sau đây thuộc về giai đoạn phân lớp trong giải thuật phân lớp k-NN?

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------------|
| A | Với mỗi ví dụ học x , tính khoảng cách giữa x và z . |
| B | Xác định tập NB (z) – các láng giềng gần nhất của z . |
| C | Phân z vào lớp chiếm số đông trong số các lớp của các ví dụ học trong NB (z). |
| D | Tính phân loại |

Đáp án đúng: C

Câu hỏi số: 188

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Tổng hợp

Mức độ: Trung bình

Phần nội dung câu hỏi:

Giải thuật dự đoán k-NN, để dự đoán giá trị đầu ra cho ví dụ z, thuộc giai đoạn nào?

Các đáp án:

| | |
|---|----------------------------|
| A | Giai đoạn học |
| B | Giai đoạn phân lớp |
| C | Giai đoạn dự đoán |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: C

Câu hỏi số: 189

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong giải thuật dự đoán k-NN, các bước thuộc giai đoạn dự đoán gồm có?

Các đáp án:

| | |
|---|--------------------------------------------------------|
| A | Đối với mỗi ví dụ học x, tính khoảng cách giữa x và z. |
| B | Xác định tập NB (z) – các láng giềng gần nhất của z. |
| C | Dự đoán giá trị đầu ra đối với z. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 190

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Tổng hợp

Mức độ: Khó

Phần nội dung câu hỏi:

Để tránh cân bằng tỉ lệ giữa các ví dụ giữa 2 lớp. Đối với bài toán phân lớp có 2 lớp, k thường được chọn như thế nào?

Các đáp án:

| | |
|---|----------------------------|
| A | Là một số chẵn |
| B | Là một số lẻ |
| C | Chọn bằng 1 |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: B

Câu hỏi số: 191

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

Có các phương pháp lựa chọn hàm khoảng cách d nào?

Các đáp án:

| | |
|---|------------------------------|
| A | Các hàm khoảng cách hình học |
| B | Các hàm khoảng cách Hamming |
| C | Hàm tính độ tương tự Cosine |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 192

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

Phương pháp lựa chọn hàm khoảng cách d nào dành cho các bài toán có các thuộc tính đầu vào là kiểu số thực?

Các đáp án:

| | |
|---|------------------------------|
| A | Các hàm khoảng cách hình học |
| B | Các hàm khoảng cách Hamming |
| C | Hàm tính độ tương tự Cosine |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: A

Câu hỏi số: 193

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

Phương pháp lựa chọn hàm khoảng cách d nào dành cho các bài toán có các thuộc tính đầu vào là kiểu nhị phân?

Các đáp án:

| | |
|---|------------------------------|
| A | Các hàm khoảng cách hình học |
| B | Các hàm khoảng cách Hamming |
| C | Hàm tính độ tương tự Cosine |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: B

Câu hỏi số: 194

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phần nội dung câu hỏi:

Phương pháp lựa chọn hàm khoảng cách d nào dành cho các bài toán phân tích văn bản?

Các đáp án:

| | |
|---|------------------------------|
| A | Các hàm khoảng cách hình học |
| B | Các hàm khoảng cách Hamming |
| C | Hàm tính độ tương tự Cosine |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: C

Câu hỏi số: 195

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong các hàm tính khoảng cách hình học (Geometry distance functions), biểu thức sau là biểu diễn của hàm gì?

$$d(x, z) = \sum_{i=1}^n |x_i - z_i|$$

Các đáp án:

| | |
|---|------------------------|
| A | Hàm Minkowski (p-norm) |
| B | Hàm Manhattan (p=1) |
| C | Hàm Euclid (p=2) |
| D | Hàm Chebyshev (p=∞) |

Đáp án đúng: B

Câu hỏi số: 196

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong các hàm tính khoảng cách hình học (Geometry distance functions), biểu thức sau là biểu diễn của hàm gì?

$$d(x, z) = \sqrt{\sum_{i=1}^n (x_i - z_i)^2}$$

Các đáp án:

| | |
|---|------------------------|
| A | Hàm Minkowski (p-norm) |
| B | Hàm Manhattan (p=1) |
| C | Hàm Euclid (p=2) |
| D | Hàm Chebyshev (p=∞) |

Đáp án đúng: C

Câu hỏi số: 197

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong các hàm tính khoảng cách hình học (Geometry distance functions), biểu thức sau biểu diễn hàm nào?

$$d(x, z) = \lim_{p \rightarrow \infty} \left(\sum_{i=1}^n |x_i - z_i|^p \right)^{1/p}$$

$$= \max_i |x_i - z_i|$$

Các đáp án:

| | |
|---|------------------------|
| A | Hàm Minkowski (p-norm) |
| B | Hàm Manhattan (p=1) |
| C | Hàm Euclid (p=2) |
| D | Hàm Chebyshev (p=∞) |

Đáp án đúng: D

Câu hỏi số: 198

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong các hàm tính khoảng cách, biểu thức sau là biểu diễn của hàm gì?

$$d(x, z) = \sum_{i=1}^n \text{Difference}(x_i, z_i)$$

$$\text{Difference}(a, b) = \begin{cases} 1, & \text{if } (a \neq b) \\ 0, & \text{if } (a = b) \end{cases}$$

Các đáp án:

| | |
|---|------------------------|
| A | Hàm Minkowski (p-norm) |
|---|------------------------|

| | |
|---|-------------------------|
| B | Hàm Manhattan (p=1) |
| C | Hàm khoảng cách Hamming |
| D | Hàm Chebyshev (p=∞) |

Đáp án đúng: C

Câu hỏi số: 199

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong các hàm tính khoảng cách, biểu thức sau là biểu diễn của hàm gì?

$$d(x, z) = \frac{x \cdot z}{\|x\| \|z\|} = \frac{\sum_{i=1}^n x_i z_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n z_i^2}}$$

Các đáp án:

| | |
|---|-----------------------------|
| A | Hàm Minkowski (p-norm) |
| B | Hàm tính độ tương tự Cosine |
| C | Hàm khoảng cách Hamming |
| D | Hàm Chebyshev (p=∞) |

Đáp án đúng: B

Câu hỏi số: 200

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong các hàm tính khoảng cách, biểu thức sau là biểu diễn của hàm gì?

$$d(x, z) = \sqrt{\sum_{i=1}^n (x_i - z_i)^2}$$

Các đáp án:

| | |
|---|-----------------------------|
| A | Hàm tính khoảng cách Euclid |
| B | Hàm tính độ tương tự Cosine |
| C | Hàm khoảng cách Hamming |

| | |
|---|------------------------------|
| D | Hàm Chebyshev ($p=\infty$) |
|---|------------------------------|

Đáp án đúng: A

Câu hỏi số: 201

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phân nội dung câu hỏi:

Biểu thức sau biểu diễn khoảng cách của các láng giềng, áp dụng đối với bài toán nào?

$$c(z) = \arg \max_{c_j \in C} \sum_{x \in NB(z)} v(x, z). \text{Identical}(c_j, c(x))$$

$$\text{Identical}(a, b) = \begin{cases} 1, & \text{if } (a = b) \\ 0, & \text{if } (a \neq b) \end{cases}$$

Các đáp án:

| | |
|---|---------------------------|
| A | Bài toán phân lớp |
| B | Bài toán dự đoán |
| C | Bài toán đệ quy |
| D | Cả 3 đáp án trên đều đúng |

Đáp án đúng: A

Câu hỏi số: 202

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phân nội dung câu hỏi:

Biểu thức sau biểu diễn khoảng cách của các láng giềng, áp dụng đối với bài toán nào?

$$f(z) = \frac{\sum_{x \in NB(z)} v(x, z).f(x)}{\sum_{x \in NB(z)} v(x, z)}$$

Các đáp án:

| | |
|---|---------------------------|
| A | Bài toán phân lớp |
| B | Bài toán dự đoán |
| C | Bài toán đệ quy |
| D | Cả 3 đáp án trên đều đúng |

Đáp án đúng: B

Câu hỏi số: 203

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phản nội dung câu hỏi:

Bài toán Lazy learning: việc đánh giá hàm mục tiêu (target function) được hoãn lại cho đến khi xét, ví dụ cần phân loại/dự đoán, gồm các đặc điểm gì?

Các đáp án:

| | |
|---|----------------------------------------------------------------------------------------|
| A | Đánh giá hàm mục tiêu một cách cụ thể và riêng rẽ cho mỗi ví dụ cần phân loại/dự đoán. |
| B | Tính toán nhiều lần các xấp xỉ cục bộ của hàm mục tiêu. |
| C | Thường mất thời gian lâu hơn để đưa ra kết luận, và cần nhiều không gian nhớ hơn. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 204

Chương: Các kỹ thuật phân lớp và dự đoán

Kỹ năng: Đánh giá

Mức độ: Trung bình

Phản nội dung câu hỏi:

Bài toán Eager learning: việc đánh giá hàm mục tiêu được hoàn thành trước khi xét đến bất kỳ, ví dụ cần phân loại/dự đoán, gồm có?

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------|
| A | Đánh giá hàm mục tiêu một cách tổng thể đối với toàn bộ không gian các ví dụ. |
| B | Tính toán một xấp xỉ duy nhất của hàm mục tiêu. |
| C | Linear regression, Support vector machines, Neural networks |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 205

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Biết

Mức độ: Trung bình

Phản nội dung câu hỏi:

Giá trị đầu ra của các nơ-ron được xác định bởi?

Các đáp án:

| | |
|---|-----------------------------------------|
| A | Đặc tính vào/ra của nó. |
| B | Các liên kết của nó với các nơ-ron khác |
| C | Các đầu vào bổ sung |
| D | Cả 3 đáp án trên đều đúng |

Đáp án đúng: D

Câu hỏi số: 206

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Biết

Mức độ: Khó

Phản nội dung câu hỏi:

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

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------------------------|
| A | Mạng nơ-ron nhân tạo (ANN: Artificial neural network) mô phỏng các hệ thống nơ-ron sinh học. |
| B | ANN là một cấu trúc (structure/network) được tạo nên bởi một số lượng nơ-ron liên kết với nhau. |
| C | Mỗi nơ-ron thực hiện một tính toán cục bộ. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 207

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Đánh giá

Mức độ: Khó

Phản nội dung câu hỏi:

Chức năng (hàm mục tiêu) của một ANN được xác định bằng gì?

Các đáp án:

| | |
|---|--------------------------------------------|
| A | Kiến trúc của mạng nơ-ron |
| B | Đặc tính vào/ra của mỗi nơ-ron |
| C | Chiến lược học (huấn luyện) và dữ liệu học |
| D | Cả 3 đáp án trên đều đúng |

Đáp án đúng: D

Câu hỏi số: 208

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Đánh giá

Mức độ: Khó

Phản nội dung câu hỏi:

Bằng cách gán và điều chỉnh các giá trị trong số của các liên kết giữa các nơ-ron, giúp cho ANN có những khả năng gì?

Các đáp án:

| | |
|---|----------------------------|
| A | Học – learn |
| B | Nhớ lại – recall |
| C | Khái quát hóa – generalize |

| | |
|---|----------------------------|
| D | Cả 3 đáp án trên đều đúng. |
|---|----------------------------|

Đáp án đúng: D

Câu hỏi số: 209

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Đánh giá

Mức độ: Dễ

Phần nội dung câu hỏi:

So khớp, tiền xử lý, phân loại và phân tích ảnh, computer vision, nén ảnh, xử lý và hiểu các ảnh thay đổi theo thời gian, thuộc loại ứng dụng nào sau đây của ANN?

Các đáp án:

| | |
|---|------------------------------|
| A | Xử lý ảnh và computer vision |
| B | Xử lý tín hiệu |
| C | Nhận dạng mẫu |
| D | Y tế |

Đáp án đúng: A

Câu hỏi số: 210

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Đánh giá

Mức độ: Khó

Phần nội dung câu hỏi:

Phân tích tín hiệu và hình thái địa chấn, động đất thuộc loại ứng dụng điển hình nào của ANN?

Các đáp án:

| | |
|---|------------------------------|
| A | Xử lý ảnh và computer vision |
| B | Xử lý tín hiệu |
| C | Nhận dạng mẫu |
| D | Y tế |

Đáp án đúng: B

Câu hỏi số: 211

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Đánh giá

Mức độ: Khó

Phần nội dung câu hỏi:

Trích chọn thuộc tính, phân loại và phân tích tín hiệu ra-đa, nhận dạng và hiểu giọng nói, nhận dạng dấu vân tay, nhận dạng chữ ký, nhận dạng mặt người, và phân tích chữ viết tay, thuộc loại ứng dụng điển hình nào của ANN?

Các đáp án:

| | |
|---|------------------------------|
| A | Xử lý ảnh và computer vision |
| B | Xử lý tín hiệu |
| C | Nhận dạng mẫu |
| D | Y tế |

Đáp án đúng: C

Câu hỏi số: 212

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Đánh giá

Mức độ: Dễ

Phản nội dung câu hỏi:

Phân tích và hiểu tín hiệu điện tim, chẩn đoán các loại bệnh, và xử lý ảnh trong lĩnh vực y tế, thuộc loại ứng dụng điển hình nào của ANN?

Các đáp án:

| | |
|---|------------------------------|
| A | Xử lý ảnh và computer vision |
| B | Xử lý tín hiệu |
| C | Nhận dạng mẫu |
| D | Y tế |

Đáp án đúng: D

Câu hỏi số: 213

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Hiểu

Mức độ: Khó

Phản nội dung câu hỏi:

Phát hiện thủy lôi, phân loại nhiễu ra-đa, thuộc loại ứng dụng điển hình nào của ANN?

Các đáp án:

| | |
|---|---------------------------------------|
| A | Các hệ thống quân sự |
| B | Các hệ thống tài chính |
| C | Lập kế hoạch, điều khiển, và tìm kiếm |
| D | Các hệ thống năng lượng |

Đáp án đúng: A

Câu hỏi số: 214

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Đánh giá

Mức độ: Khó

Phản nội dung câu hỏi:

Phân tích thị trường chứng khoán, đánh giá giá trị bất động sản, kiểm tra truy cập thẻ tín dụng, kinh doanh cổ phiếu, đều thuộc loại ứng dụng điển hình nào của ANN?

Các đáp án:

| | |
|---|---------------------------------------|
| A | Các hệ thống quân sự |
| B | Các hệ thống tài chính |
| C | Lập kế hoạch, điều khiển, và tìm kiếm |
| D | Các hệ thống năng lượng |

Đáp án đúng: B

Câu hỏi số: 215

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Tổng hợp

Mức độ: Khó

Phần nội dung câu hỏi:

Cài đặt song song các bài toán thỏa mãn ràng buộc, tìm lời giải cho bài toán người đưa hàng, điều khiển và khoa học nghiên cứu về người máy, đều thuộc loại ứng dụng điển hình nào của ANN?

Các đáp án:

| | |
|---|---------------------------------------|
| A | Các hệ thống quân sự |
| B | Các hệ thống tài chính |
| C | Lập kế hoạch, điều khiển, và tìm kiếm |
| D | Các hệ thống năng lượng |

Đáp án đúng: B

Câu hỏi số: 216

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Biết

Mức độ: Dễ

Phần nội dung câu hỏi:

Đánh giá trạng thái hệ thống, phát hiện và khắc phục sự cố, dự đoán tải (khối lượng) công việc, và đánh giá mức độ an toàn, đều thuộc loại ứng dụng điển hình nào của ANN ?

Các đáp án:

| | |
|---|---------------------------------------|
| A | Các hệ thống quân sự |
| B | Các hệ thống tài chính |
| C | Lập kế hoạch, điều khiển, và tìm kiếm |
| D | Các hệ thống năng lượng |

Đáp án đúng: D

Câu hỏi số: 217

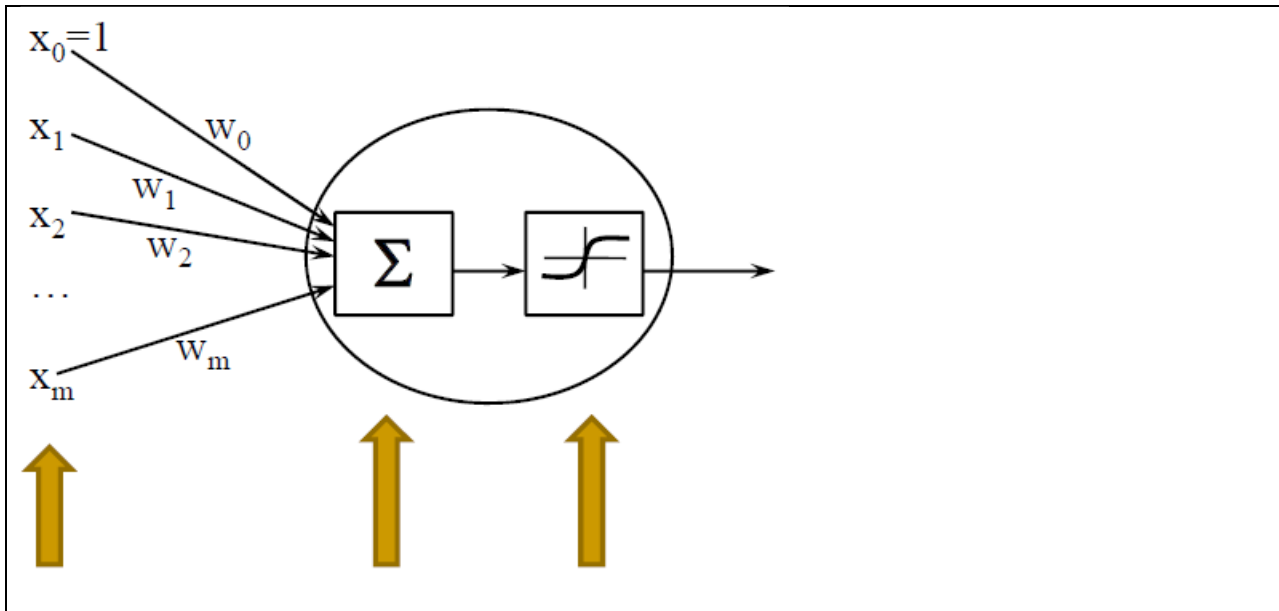
Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Biết

Mức độ: Dễ

Phần nội dung câu hỏi:

Cấu trúc của một nơ-ron bao gồm những thành phần nào?



Các đáp án:

| | |
|---|----------------------------|
| A | Các tín hiệu đầu vào |
| B | Hàm tác động/truyền |
| C | Giá trị đầu ra |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 218

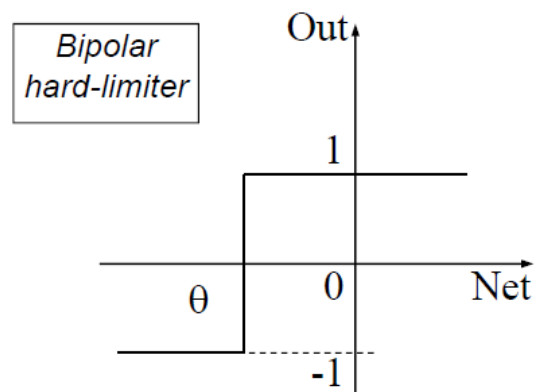
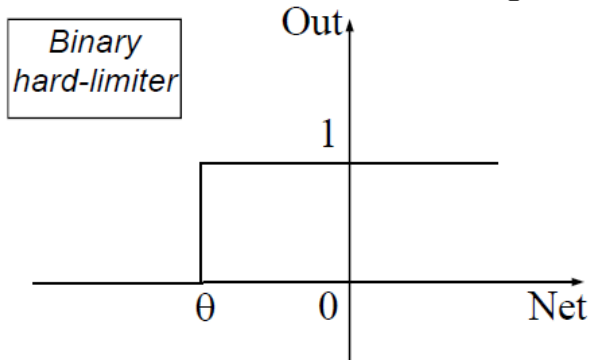
Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Biết

Mức độ: Dễ

Phản nội dung câu hỏi:

Đồ thị sau biểu diễn loại hàm tác động nào?



Các đáp án:

| | |
|---|--------------------|
| A | Giới hạn chặn |
| B | Logic ngưỡng |
| C | Xích-ma |
| D | Hyperbolic tangent |

Đáp án đúng: D

Câu hỏi số: 219

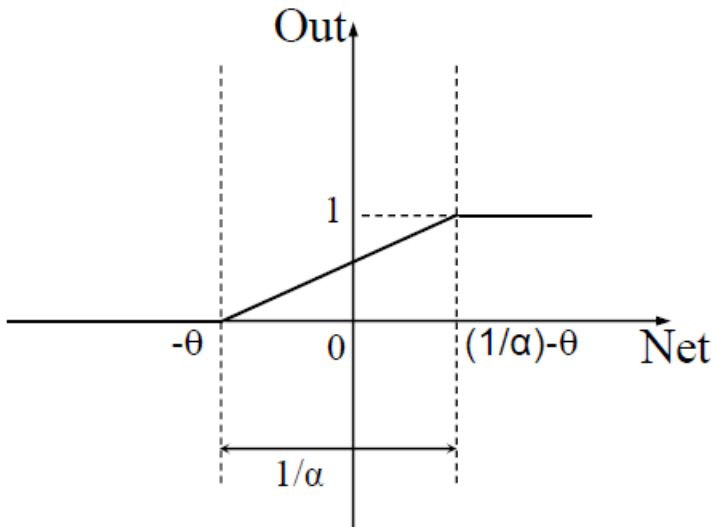
Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Biết

Mức độ: Dễ

Phản nội dung câu hỏi:

Đồ thị sau biểu diễn loại hàm tác động nào?



Các đáp án:

| | |
|---|--------------------|
| A | Giới hạn chặn |
| B | Logic ngưỡng |
| C | Xích-ma |
| D | Hyperbolic tangent |

Đáp án đúng: B

Câu hỏi số: 220

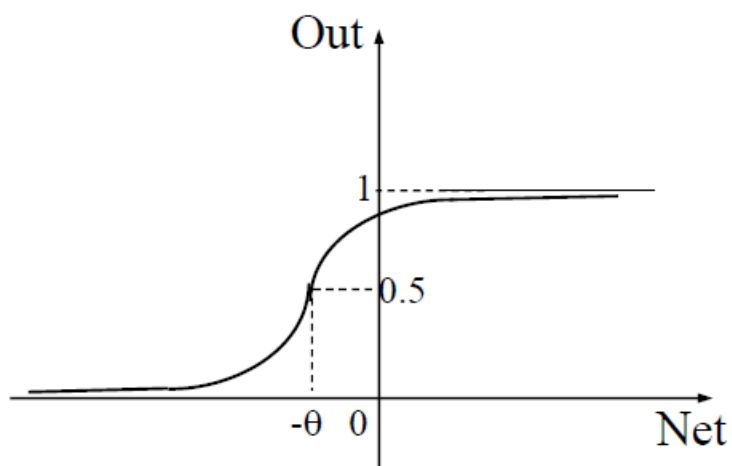
Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Biết

Mức độ: Dễ

Phản nội dung câu hỏi:

Đồ thị sau biểu diễn loại hàm tác động nào?



Các đáp án:

| | |
|---|--------------------|
| A | Giới hạn chặn |
| B | Logic ngưỡng |
| C | Xích-ma |
| D | Hyperbolic tangent |

Đáp án đúng: C

Câu hỏi số: 221

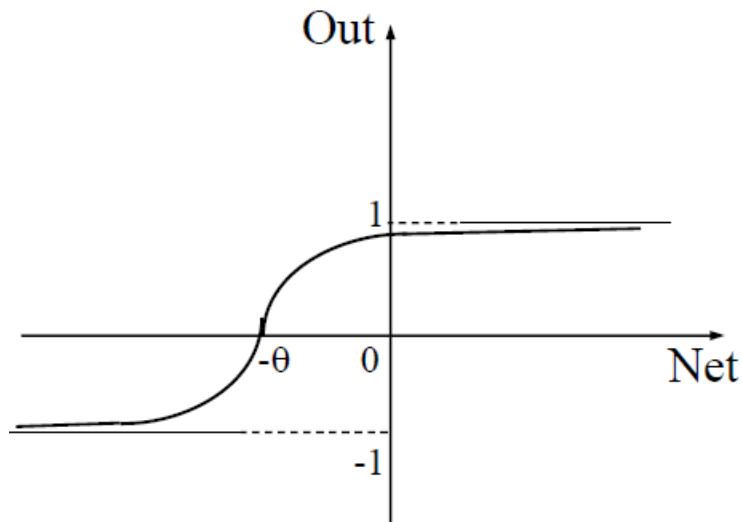
Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Biết

Mức độ: Dễ

Phản nội dung câu hỏi:

Đồ thị sau biểu diễn loại hàm tác động nào?



Các đáp án:

| | |
|---|--------------------|
| A | Giới hạn chặn |
| B | Logic ngưỡng |
| C | Xích-ma |
| D | Hyperbolic tangent |

Đáp án đúng: D

Câu hỏi số: 222

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Biết

Mức độ: Dễ

Phản nội dung câu hỏi:

Trong các hàm tác động, hàm nào sau đây không liên tục và đạo hàm của nó cũng không liên tục?

Các đáp án:

| | |
|---|--------------------|
| A | Giới hạn chặn |
| B | Logic ngưỡng |
| C | Xích-ma |
| D | Hyperbolic tangent |

Đáp án đúng: A

Câu hỏi số: 223

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Biết

Mức độ: Dễ

Phản nội dung câu hỏi:

Hàm tác động nào sau đây là liên tục nhưng đạo hàm của nó lại không liên tục?

Các đáp án:

| | |
|---|--------------------|
| A | Giới hạn chặn |
| B | Logic ngưỡng |
| C | Xích-ma |
| D | Hyperbolic tangent |

Đáp án đúng: B

Câu hỏi số: 224

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Biết

Mức độ: Dễ

Phần nội dung câu hỏi:

Hàm tác động nào sau đây là liên tục; và đạo hàm của nó là liên tục và được biểu diễn bằng một hàm của chính nó?

Các đáp án:

| | |
|---|--------------------|
| A | Giới hạn chặn |
| B | Logic ngưỡng |
| C | Xích-ma |
| D | Hyperbolic tangent |

Đáp án đúng: C

Câu hỏi số: 225

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Hiểu

Mức độ: Dễ

Phần nội dung câu hỏi:

Hàm tác động nào liên tục và yêu cầu tham số anpha xác định độ dốc, với giá trị đầu ra trong khoảng $(-1,1)$?

Các đáp án:

| | |
|---|--------------------|
| A | Giới hạn chặn |
| B | Logic ngưỡng |
| C | Xích-ma |
| D | Hyperbolic tangent |

Đáp án đúng: D

Câu hỏi số: 226

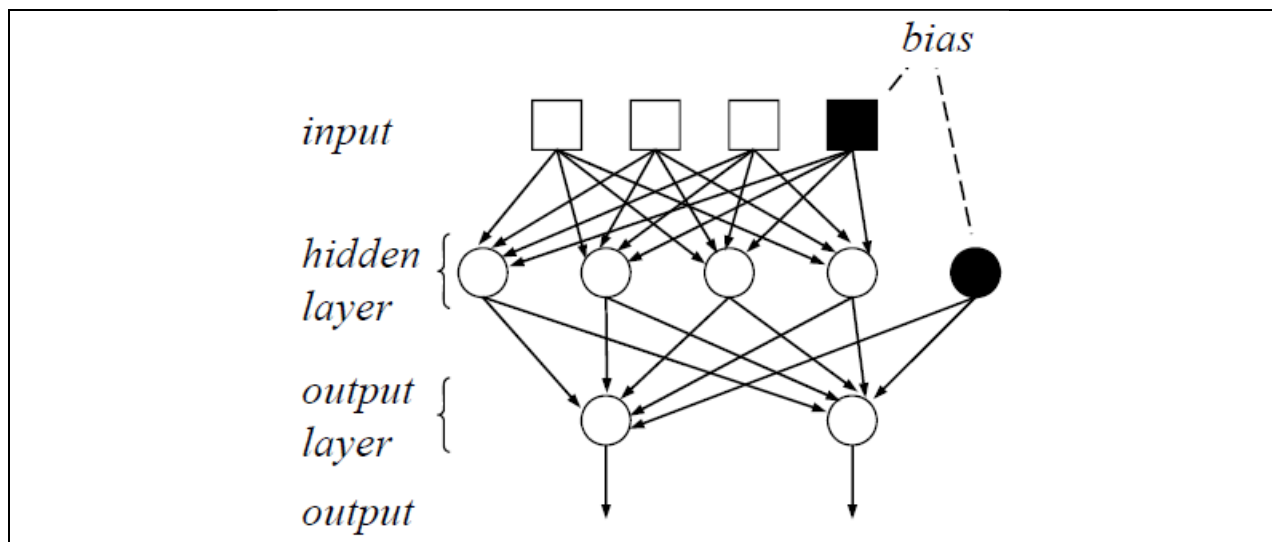
Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

Một ANN phải có những thành phần nào?



Các đáp án:

| | |
|---|--------------------------------|
| A | Một tầng đầu vào |
| B | Một tầng đầu ra |
| C | Không, một, hoặc nhiều tầng ẩn |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 227

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

Kiến trúc của một ANN được xác định bởi những yếu tố nào?

Các đáp án:

| | |
|---|--------------------------------------------------------|
| A | Số lượng các tín hiệu đầu vào và đầu ra |
| B | Số lượng các tầng, số lượng các nơ-ron trong mỗi tầng. |
| C | Số lượng các trọng số đối với mỗi nơ-ron |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 228

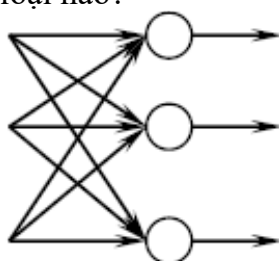
Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

Kiến trúc mạng Nơ-ron sau thuộc loại nào?



Các đáp án:

| | |
|---|-------------------------------|
| A | Mạng lan truyền tiến một tầng |
|---|-------------------------------|

| | |
|---|---------------------------------|
| B | Mạng lan truyền tiến nhiều tầng |
| C | Mạng hồi qui một tầng |
| D | Mạng hồi qui nhiều tầng |

Đáp án đúng: A

Câu hỏi số: 229

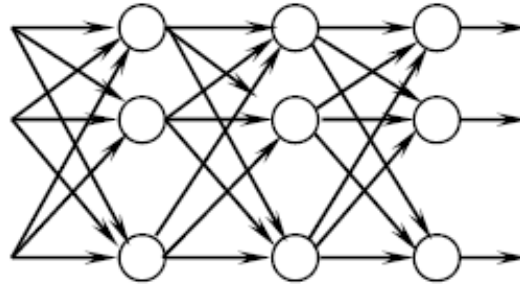
Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Hiểu

Mức độ: Trung bình

Phản nội dung câu hỏi:

Kiến trúc mạng Nơ-ron sau thuộc loại nào?



Các đáp án:

| | |
|---|---------------------------------|
| A | Mạng lan truyền tiến một tầng |
| B | Mạng lan truyền tiến nhiều tầng |
| C | Mạng hồi qui một tầng |
| D | Mạng hồi qui nhiều tầng |

Đáp án đúng: B

Câu hỏi số: 230

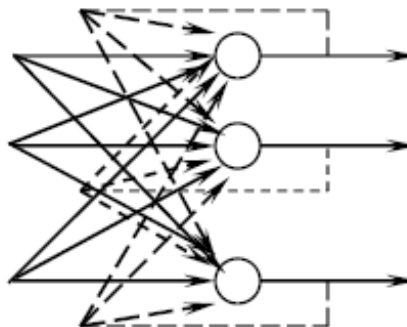
Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Hiểu

Mức độ: Trung bình

Phản nội dung câu hỏi:

Kiến trúc mạng Nơ-ron sau thuộc loại nào?



Các đáp án:

| | |
|---|---------------------------------|
| A | Mạng lan truyền tiến một tầng |
| B | Mạng lan truyền tiến nhiều tầng |
| C | Mạng hồi qui một tầng |
| D | Mạng hồi qui nhiều tầng |

Đáp án đúng: C

Câu hỏi số: 231

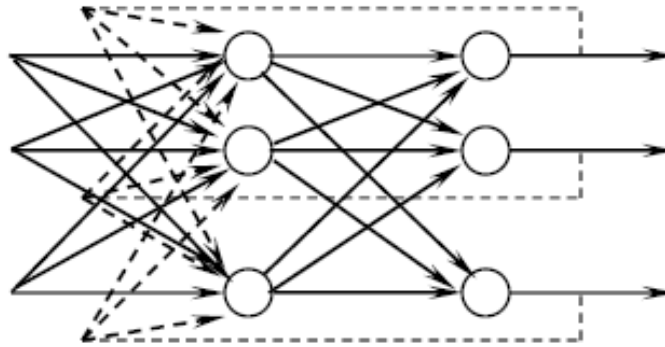
Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

Kiến trúc mạng Nơ-ron sau thuộc loại nào?



Các đáp án:

| | |
|---|---------------------------------|
| A | Mạng lan truyền tiến một tầng |
| B | Mạng lan truyền tiến nhiều tầng |
| C | Mạng hồi qui một tầng |
| D | Mạng hồi qui nhiều tầng |

Đáp án đúng: D

Câu hỏi số: 232

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

Phát biểu nào sau đây là đúng về giải thuật học của Perception?

Các đáp án:

| | |
|---|----------------------------------------------------------------------------------------------------------------------------------------------------|
| A | Quá trình học của perception nhằm xác định một véc-tơ trọng số cho phép perceptron sinh ra giá trị đầu ra chính xác (-1 hoặc 1) cho mỗi ví dụ học. |
| B | Nếu $d=1$ nhưng perceptron lại sinh ra -1 ($Out=-1$), thì w cần được thay đổi sao cho giá trị $Net(w,x)$ tăng lên. |
| C | Nếu $d=-1$ nhưng perceptron lại sinh ra 1 ($Out=1$), thì w cần được thay đổi sao cho giá trị $Net(w,x)$ giảm đi. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 233

Chương: Mạng nơ-ron nhân tạo

Kỹ năng: Hiểu

Mức độ: Khó

Phần nội dung câu hỏi:

Trong giải thuật học lan truyền ngược cần quan tâm đến những vấn đề gì?

Các đáp án:

| | |
|---|------------------------------------------------------------------------------------|
| A | Căn cứ vào giá trị đầu ra mong muốn của véc-tơ đầu vào, hệ thống tính toán giá trị |
|---|------------------------------------------------------------------------------------|

| | |
|---|------------------------------------------------------------------------------------------------------------------------|
| | lỗi. |
| B | Bắt đầu từ tầng đầu ra, giá trị lỗi được lan truyền ngược qua mạng, từ tầng này qua tầng khác, cho đến tầng cuối cùng. |
| C | Việc lan truyền ngược lỗi được thực hiện thông qua việc tính toán giá trị gradient cục bộ của mỗi nơ-ron. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 234

Chương: Các kỹ thuật phân nhóm

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phân nội dung câu hỏi:

| |
|-------------------------------------------------------|
| Học có giám sát (Supervised learning) có đặc điểm gì? |
|-------------------------------------------------------|

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------------------------------------|
| A | Tập dữ liệu (dataset) bao gồm các ví dụ, mà mỗi ví dụ được gắn kèm với một nhãn lớp/giá trị đầu ra mong muốn. |
| B | Mục đích là học (xấp xỉ) một giả thiết phù hợp với tập dữ liệu hiện có. |
| C | Giả thiết học được (learned hypothesis) sau đó sẽ được dùng để phân lớp/dự đoán đối với các ví dụ mới. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 235

Chương: Các kỹ thuật phân nhóm

Kỹ năng: Hiểu

Mức độ: Trung bình

Phân nội dung câu hỏi:

| |
|------------------------------------------------------------|
| Học không giám sát (Unsupervised learning) có đặc điểm gì? |
|------------------------------------------------------------|

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------------------------------------|
| A | Tập dữ liệu (dataset) bao gồm các ví dụ, mà mỗi ví dụ được gắn kèm với một nhãn lớp/giá trị đầu ra mong muốn. |
| B | Mục đích là học (xấp xỉ) một giả thiết phù hợp với tập dữ liệu hiện có. |
| C | Giả thiết học được (learned hypothesis) sau đó sẽ được dùng để phân lớp/dự đoán đối với các ví dụ mới. |
| D | Tập dữ liệu bao gồm các ví dụ, mà mỗi ví dụ không có thông tin về nhãn lớp/giá trị đầu ra mong muốn. |

Đáp án đúng: D

Câu hỏi số: 236

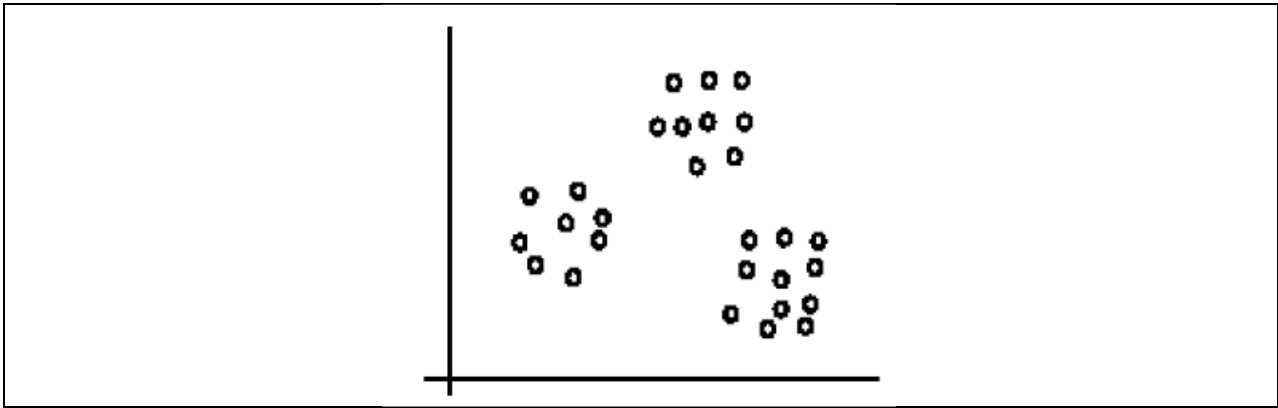
Chương: Các kỹ thuật phân nhóm

Kỹ năng: Áp dụng

Mức độ: Dễ

Phân nội dung câu hỏi:

| |
|------------------------------------------------------------------------|
| Ví dụ sau về phân nhóm – trong đó, các ví dụ được chia thành mấy nhóm? |
|------------------------------------------------------------------------|



Các đáp án:

| | |
|---|---|
| A | 2 |
| B | 3 |
| C | 4 |
| D | 5 |

Đáp án đúng: B

Câu hỏi số: 237

Chương: Các kỹ thuật phân nhóm

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phản nội dung câu hỏi:

| |
|-----------------------------------------------------------------------|
| Giải thuật K-Means phân chia (partitions) tập dữ liệu thành mấy nhóm? |
|-----------------------------------------------------------------------|

Các đáp án:

| | |
|---|-------------------|
| A | k |
| B | K^{mean} |
| C | $K!$ |
| D | 2k |

Đáp án đúng: A

Câu hỏi số: 238

Chương: Các kỹ thuật phân nhóm

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phản nội dung câu hỏi:

| |
|---------------------------------------|
| Quá trình phân nhóm kết thúc khi nào? |
|---------------------------------------|

Các đáp án:

| | |
|---|------------------------------------------------------------------------------|
| A | Không có (hoặc có không đáng kể) việc gán lại các ví dụ vào các nhóm khác |
| B | Không có (hoặc có không đáng kể) thay đổi về các điểm trung tâm của các nhóm |
| C | Giảm không đáng kể về tổng lỗi phân nhóm |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 239

Chương: Các kỹ thuật phân nhóm

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phản nội dung câu hỏi:

Các ưu điểm của giải thuật K-Means gồm có?

Các đáp án:

| | |
|---|----------------------------------------------------------------------------------------------------------|
| A | Đơn giản |
| B | Nếu cả 2 giá trị k và t đều nhỏ, thì giải thuật k-means được xem như là có độ phức tạp ở mức tuyến tính. |
| C | Được dùng phổ biến |
| D | Cả 3 đáp án trên đều đúng |

Đáp án đúng: D

Câu hỏi số: 240

Chương: Các kỹ thuật phân nhóm

Kỹ năng: Hiểu

Mức độ: Trung bình

Phản nội dung câu hỏi:

Nhược điểm của giải thuật K-Means là gì?

Các đáp án:

| | |
|---|-------------------------------------------------------|
| A | Giá trị k (số nhóm thu được) phải được xác định trước |
| B | Cần xác định cách tính điểm trung bình của một nhóm |
| C | Dễ gặp lỗi với các ví dụ ngoại lai |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 241

Chương: Các kỹ thuật phân nhóm

Kỹ năng: Biết

Mức độ: Dễ

Phản nội dung câu hỏi:

Trong giải thuật k-means, các ngoại lai (outliers) là các ví dụ như thế nào?

Các đáp án:

| | |
|---|--------------------------------------------------------------------------------------------------|
| A | Các ví dụ rất khác biệt với tất cả các ví dụ khác |
| B | Các ví dụ lỗi trong quá trình thu thập/lưu dữ liệu |
| C | Các ví dụ có các giá trị thuộc tính rất khác biệt với các giá trị thuộc tính của các ví dụ khác. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 242

Chương: Các kỹ thuật phân nhóm

Kỹ năng: Biết

Mức độ: Dễ

Phản nội dung câu hỏi:

Giải pháp nào sau đây giải quyết được vấn đề ngoại lai?

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------------------------------------------|
| A | Trong quá trình phân nhóm, cần loại bỏ một số các ví dụ quá khác biệt với các điểm trung tâm so với các ví dụ khác. |
| B | Thực hiện việc lấy mẫu theo tuần tự. |
| C | Gán các ví dụ khác biệt của tập dữ liệu vào các nhóm tùy theo đánh giá về khoảng |

| | |
|---|-----------------------------------------------------------------------------------------------------|
| | cách. |
| D | Trong quá trình phân nhóm, cần thêm một số các ví dụ giống với các điểm trung tâm ở các ví dụ khác. |

Đáp án đúng: A

Câu hỏi số: 243

Chương: Các kỹ thuật phân nhóm

Kỹ năng: Biết

Mức độ: Dễ

Phân nội dung câu hỏi:

Do quá trình lấy mẫu chỉ lựa chọn một tập con nhỏ của tập dữ liệu ban đầu, nên khả năng một ngoại lai được chọn là rất nhỏ. Giải pháp nào đảm bảo điều này?

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------------------------------------------|
| A | Trong quá trình phân nhóm, cần loại bỏ một số các ví dụ quá khác biệt với các điểm trung tâm so với các ví dụ khác. |
| B | Thực hiện việc lấy mẫu theo tuần tự. |
| C | Thực hiện việc lấy mẫu ngẫu nhiên. |
| D | Trong quá trình phân nhóm, cần thêm một số các ví dụ giống với các điểm trung tâm ở các ví dụ khác. |

Đáp án đúng: C

Câu hỏi số: 244

Chương: Các kỹ thuật phân nhóm

Kỹ năng: Biết

Mức độ: Dễ

Phân nội dung câu hỏi:

Hình sau thể hiện nhược điểm gì của giải thuật k-means?

(A). Random selection of seeds (centroids)

(B). Iteration 1

(C). Iteration 2

Các đáp án:

| | |
|---|----------------------------------------------------------------------------------|
| A | Sử dụng các hạt nhân khác nhau. |
| B | Phụ thuộc việc chọn các điểm trung tâm ban đầu. |
| C | Không phù hợp để phát hiện các nhóm (cụm) không có dạng hình elip hoặc hình cầu. |

| | |
|---|----------------------------|
| D | Cả 3 đáp án trên đều đúng. |
|---|----------------------------|

Đáp án đúng: B

Câu hỏi số: 245

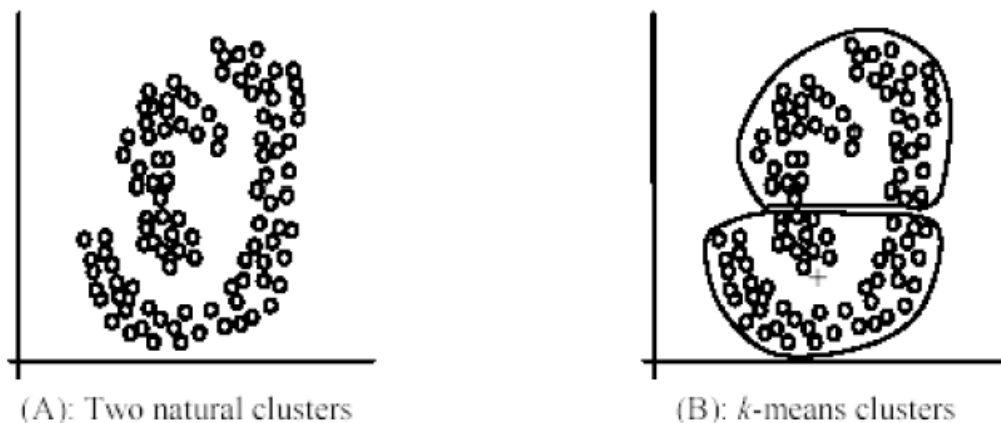
Chương: Các kỹ thuật phân nhóm

Kỹ năng: Hiểu

Mức độ: Dễ

Phân nội dung câu hỏi:

Hình sau thể hiện nhược điểm gì của giải thuật k-means?



Các đáp án:

| | |
|---|----------------------------------------------------------------------------------|
| A | Sử dụng các hạt nhân khác nhau. |
| B | Phụ thuộc việc chọn các điểm trung tâm ban đầu. |
| C | Không phù hợp để phát hiện các nhóm (cụm) không có dạng hình elip hoặc hình cầu. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: C

Câu hỏi số: 246

Chương: Các kỹ thuật phân nhóm

Kỹ năng: Biết

Mức độ: Dễ

Phân nội dung câu hỏi:

Phương pháp nào sau đây dùng để đánh giá khoảng cách giữa 2 nhóm – đưa đến các biến thể khác nhau của giải thuật HAC?

Các đáp án:

| | |
|---|------------------------------------|
| A | Liên kết đơn (single link) |
| B | Liên kết hoàn toàn (Complete link) |
| C | Liên kết trung bình (Average link) |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 247

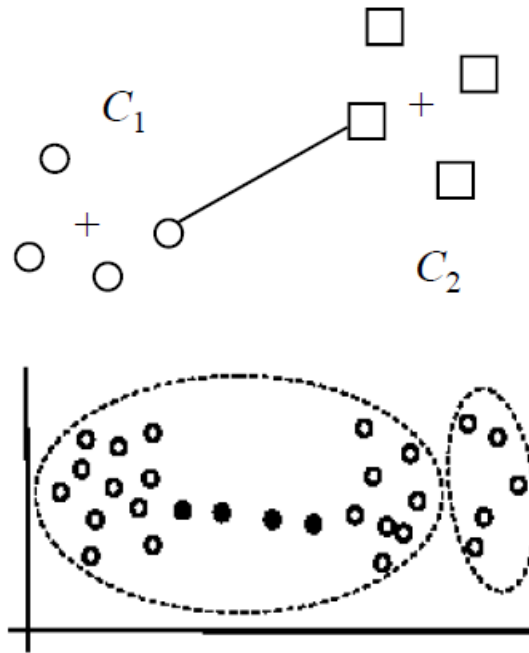
Chương: Các kỹ thuật phân nhóm

Kỹ năng: Biết

Mức độ: Dễ

Phân nội dung câu hỏi:

Hình sau thể hiện liên kết gì trong giải thuật HAC?



Các đáp án:

| | |
|---|------------------------------------|
| A | Liên kết đơn (single link) |
| B | Liên kết hoàn toàn (Complete link) |
| C | Liên kết trung bình (Average link) |
| D | Liên kết trung tâm (Centroid link) |

Đáp án đúng: A

Câu hỏi số: 248

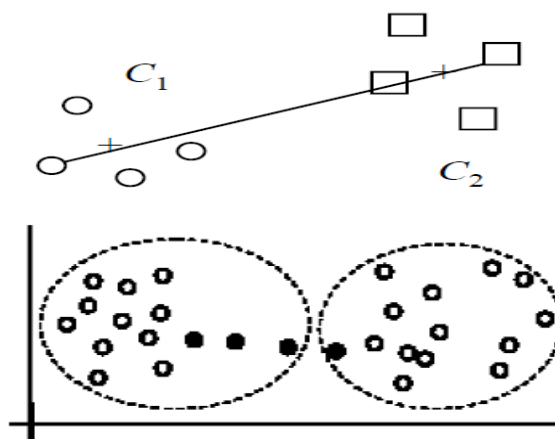
Chương: Các kỹ thuật phân nhóm

Kỹ năng: Biết

Mức độ: Dễ

Phản nội dung câu hỏi:

Hình sau thể hiện liên kết gì trong giải thuật HAC?



Các đáp án:

| | |
|---|------------------------------------|
| A | Liên kết đơn (single link) |
| B | Liên kết hoàn toàn (Complete link) |
| C | Liên kết trung bình (Average link) |
| D | Liên kết trung tâm (Centroid link) |

Đáp án đúng: B

Câu hỏi số: 249

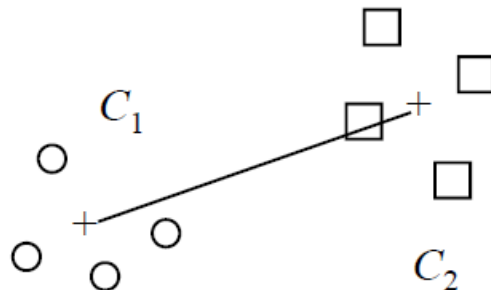
Chương: Các kỹ thuật phân nhóm

Kỹ năng: Biết

Mức độ: Dễ

Phần nội dung câu hỏi:

Hình sau thể hiện liên kết nào trong giải thuật HAC?



Các đáp án:

| | |
|---|------------------------------------|
| A | Liên kết đơn (single link) |
| B | Liên kết hoàn toàn (Complete link) |
| C | Liên kết trung bình (Average link) |
| D | Liên kết trung tâm (Centroid link) |

Đáp án đúng: D

Câu hỏi số: 250

Chương: Các kỹ thuật phân nhóm

Kỹ năng: Áp dụng

Mức độ: Dễ

Phần nội dung câu hỏi:

Hàm khoảng cách cho thuộc tính số thường dùng là gì?

Các đáp án:

| | |
|---|---------------------------------------------------------|
| A | Họ các hàm khoảng cách hình học (khoảng cách Minkowski) |
| B | Khoảng cách Euclid |
| C | Khoảng cách Manhattan |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 251

Chương: Tổng hợp

Kỹ năng: Áp dụng

Mức độ: Dễ

Phần nội dung câu hỏi:

Về phía người sử dụng, các khó khăn gặp phải với CSDL là gì?

Các đáp án:

| | |
|---|-----------------------------------------|
| A | Không thể tìm thấy dữ liệu cần thiết |
| B | Không thể lấy ra được dữ liệu cần thiết |
| C | Không thể hiểu dữ liệu tìm thấy |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 252

Chương: Tổng hợp

Kỹ năng: Áp dụng

Mức độ: Dễ

Phản nội dung câu hỏi:

Những vấn đề về hệ thống thông tin đối với hệ CSDL là gì?

Các đáp án:

| | |
|---|------------------------------------------------------------------|
| A | Phát triển các chương trình ứng dụng khác nhau là không đơn giản |
| B | Duy trì các chương trình ứng dụng gặp rất nhiều vấn đề |
| C | Khối lượng dữ liệu lưu trữ tăng rất nhanh |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 253

Chương: Tổng hợp

Kỹ năng: Áp dụng

Mức độ: Dễ

Phản nội dung câu hỏi:

Ứng dụng nào sau đây có thể được sử dụng trong Data Mining?

Các đáp án:

| | |
|---|----------------------------|
| A | Microsoft Visual Studio |
| B | Data Analysis |
| C | NorthWind |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 254

Chương: Tổng hợp

Kỹ năng: Đánh giá

Mức độ: Dễ

Phản nội dung câu hỏi:

Quy trình nào sau đây là quy trình phát hiện tri thức?

Các đáp án:

| | |
|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A | Thu thập và tiền xử lý dữ liệu -> Hình thành, xác định và định nghĩa bài toán -> Khai phá dữ liệu, rút ra các tri thức -> Sử dụng các tri thức phát hiện được. |
| B | Hình thành, xác định và định nghĩa bài toán -> Thu thập và tiền xử lý dữ liệu -> Khai phá dữ liệu, rút ra các tri thức -> Sử dụng các tri thức phát hiện được. |
| C | Hình thành, xác định và định nghĩa bài toán -> Khai phá dữ liệu, rút ra các tri thức -> Thu thập và tiền xử lý dữ liệu -> Sử dụng các tri thức phát hiện được. |
| D | Hình thành, xác định và định nghĩa bài toán -> Thu thập và tiền xử lý dữ liệu -> Sử dụng các tri thức phát hiện được -> Khai phá dữ liệu, rút ra các tri thức. |

Đáp án đúng: B

Câu hỏi số: 255

Chương: Tổng hợp

Kỹ năng: Phân tích

Mức độ: Dễ

Phần nội dung câu hỏi:

Nhiệm vụ chính trong khai phá dữ liệu là gì?

Các đáp án:

| | |
|---|----------------------------|
| A | Phân lớp |
| B | Hồi qui |
| C | Tổng hợp |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 256

Chương: Tổng hợp

Kỹ năng: Áp dụng

Mức độ: Dễ

Phần nội dung câu hỏi:

Nhiệm vụ nào tập trung vào khám phá hầu hết sự thay đổi có nghĩa dưới dạng độ đo đã biết trước hoặc giá trị chuẩn, phát hiện độ lệch đáng kể giữa nội dung của tập con dữ liệu thực và nội dung mong đợi?

Các đáp án:

| | |
|---|----------------------------------|
| A | Phân cụm |
| B | Tổng hợp |
| C | Mô hình hóa sự phụ thuộc |
| D | Phát hiện sự biến đổi và độ lệch |

Đáp án đúng: D

Câu hỏi số: 257

Chương: Tổng hợp

Kỹ năng: Biết

Mức độ: Dễ

Phần nội dung câu hỏi:

Giải thuật khai phá dữ liệu bao gồm những thành phần chính nào sau đây?

Các đáp án:

| | |
|---|----------------------------|
| A | Biểu diễn mô hình |
| B | Kiểm định mô hình |
| C | Phương pháp tìm kiếm |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 258

Chương: Tổng hợp

Kỹ năng: Biết

Mức độ: Dễ

Phần nội dung câu hỏi:

Phương pháp tự tìm kiếm, tạo mẫu và sinh ra tri thức chứ không phải bắt đầu với tri thức đã biết trước, còn được gọi là gì?

Các đáp án:

| | |
|---|---------------------------------------|
| A | Phương pháp suy diễn |
| B | Phương pháp quy nạp |
| C | Phương pháp ứng dụng K-láng giềng gần |

| | |
|---|--------------------------------------------|
| D | Phương pháp sử dụng cây quyết định và luật |
|---|--------------------------------------------|

Đáp án đúng: B

Câu hỏi số: 259

Chương: Tổng hợp

Kỹ năng: Biết

Mức độ: Dễ

Phần nội dung câu hỏi:

| |
|----------------------------------------------------------------------------------------------------|
| Nhằm rút ra thông tin là kết quả logic của các thông tin trong CSDL, phương pháp nào được áp dụng? |
|----------------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|--------------------------------------------|
| A | Phương pháp suy diễn |
| B | Phương pháp quy nạp |
| C | Phương pháp ứng dụng K-láng giềng gần |
| D | Phương pháp sử dụng cây quyết định và luật |

Đáp án đúng: A

Câu hỏi số: 260

Chương: Tổng hợp

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

| |
|-------------------------------------------------------------|
| Khai phá dữ liệu rất khác với phát kiến khoa học ở chỗ nào? |
|-------------------------------------------------------------|

Các đáp án:

| | |
|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A | Khai phá trong CSDL ít có chủ tâm và có điều kiện hơn. |
| B | Dữ liệu khoa học có từ thực nghiệm nhằm loại bỏ một số tác động của các tham số để nhấn mạnh độ biến thiên của một hay một số tham số đích. Trong khi các CSDL thương mại lại ghi một số lượng thừa thông tin. |
| C | Các nhà khoa học có thể tạo lại các thí nghiệm, trong khi các nhà quản lý CSDL hầu như không thiết kế lại các trường dữ liệu và thu thập lại dữ liệu. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 261

Chương: Tổng hợp

Kỹ năng: Áp dụng

Mức độ: Khó

Phần nội dung câu hỏi:

| |
|-------------------------------------------------------|
| Khai phá dữ liệu khác gì so với phương pháp thống kê? |
|-------------------------------------------------------|

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------------------------------------|
| A | Các phương pháp thống kê chuẩn không phù hợp đối với các kiểu dữ liệu có cấu trúc trong rất nhiều các CSDL. |
| B | Thống kê hoàn toàn theo dữ liệu (data driven), nó không sử dụng tri thức sẵn có về lĩnh vực. |
| C | Các kết quả phân tích thống kê có thể sẽ rất nhiều và khó có thể làm rõ được. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 262

Chương: Tổng hợp

Kỹ năng: Biết

Mức độ: Dễ

Phân nội dung câu hỏi:

Trong rất nhiều ứng dụng, điều quan trọng là những điều khai thác được phải càng dễ hiểu với con người càng tốt. Điều này được thể hiện ở đâu?

Các đáp án:

| | |
|---|-------------------------------------------------------|
| A | Đánh giá tầm quan trọng thống kê |
| B | Khả năng biểu đạt của mẫu |
| C | Sự tương tác với người sử dụng và các tri thức sẵn có |
| D | Quá phù hợp |

Đáp án đúng: B

Câu hỏi số: 263

Chương: Tổng hợp

Kỹ năng: Biết

Mức độ: Dễ

Phân nội dung câu hỏi:

Giải thuật tìm kiếm tham số tốt nhất sử dụng một tập dữ liệu hữu hạn có thể bị tình trạng “quá độ”, làm cho mô hình hoạt động rất kém đối với các dữ liệu thử. Vấn đề này là gì?

Các đáp án:

| | |
|---|-------------------------------------------------------|
| A | Đánh giá tầm quan trọng thống kê |
| B | Khả năng biểu đạt của mẫu |
| C | Sự tương tác với người sử dụng và các tri thức sẵn có |
| D | Quá phù hợp |

Đáp án đúng: D

Câu hỏi số: 264

Chương: Tổng hợp

Kỹ năng: Biết

Mức độ: Khó

Phân nội dung câu hỏi:

Các giải pháp như đưa ra một ngưỡng cho CSDL, lấy mẫu, các phương pháp xấp xỉ, xử lý son song,... nhằm giải quyết vấn đề gì?

Các đáp án:

| | |
|---|-----------------------------------|
| A | Độ nhiễu và tính không chắc chắn. |
| B | Các giá trị của CSDL bị thiếu |
| C | Dữ liệu lớn |
| D | Dữ liệu động |

Đáp án đúng: C

Câu hỏi số: 265

Chương: Tổng hợp

Kỹ năng: Đánh giá

Mức độ: Dễ

Phân nội dung câu hỏi:

Chuẩn hóa là một phần hữu ích của thuật toán phân lớp trong mạng nơ-ron, hoặc thuật toán tính toán độ lệch sử dụng trong việc phân lớp hay nhóm cụm các phần tử liên kề. Có các phương pháp nào được xem xét?

Các đáp án:

| | |
|---|----------------------------|
| A | Min-Max |
| B | z-score |
| C | Thay đổi chữ số thập phân |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 266

Chương: Tổng hợp

Kỹ năng: Phân tích

Mức độ: Trung bình

Phần nội dung câu hỏi:

Các kỹ thuật lựa chọn thường dùng là gì?

Các đáp án:

| | |
|---|----------------------------|
| A | Lựa chọn tăng dần |
| B | Loại bớt |
| C | Cây quyết định |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 267

Chương: Tổng hợp

Kỹ năng: Ứng dụng

Mức độ: Dễ

Phần nội dung câu hỏi:

Một số hướng tiếp cận trong khai phá luật kết hợp gồm có?

Các đáp án:

| | |
|---|------------------------------------------------------------------------------|
| A | Luật kết hợp nhị phân; Luật kết hợp có thuộc tính số và thuộc tính hạng mục. |
| B | Luật kết hợp tiếp cận theo hướng tập thô; Luật kết hợp nhiều mức. |
| C | Luật kết hợp mờ; Luật kết hợp với thuộc tính được đánh trọng số. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 268

Chương: Tổng hợp

Kỹ năng: Ứng dụng

Mức độ: Dễ

Phần nội dung câu hỏi:

Các ưu điểm của áp dụng tập mờ để rời rạc hóa tín hiệu gồm có?

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------------|
| A | Cho phép biểu diễn luật kết hợp dưới dạng tự nhiên và gần gũi hơn với người sử dụng |
| B | Số lượng tập mờ gắn với mỗi thuộc tính là không đáng kể. |
| C | Giải quyết được vấn đề “điểm biên gãy”. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 269

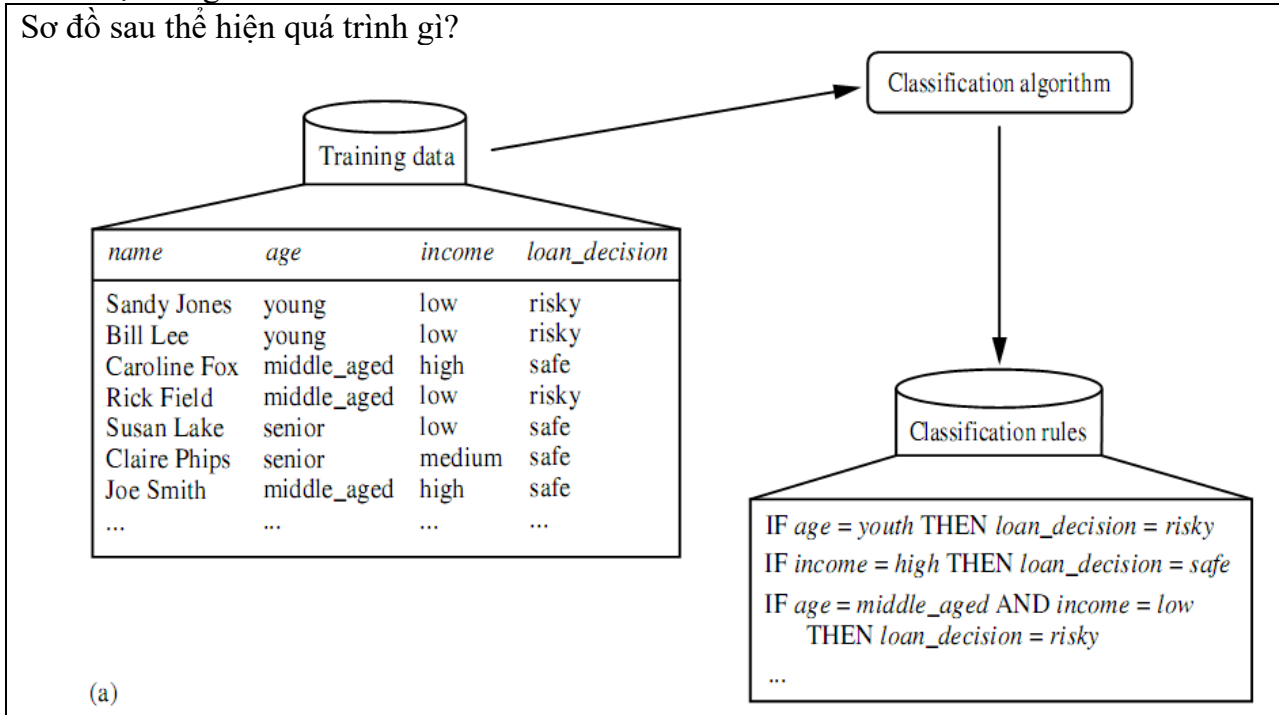
Chương: Tổng hợp

Kỹ năng: Phân tích

Mức độ: Trung bình

Phân nội dung câu hỏi:

Sơ đồ sau thể hiện quá trình gì?



Các đáp án:

| | |
|---|--------------------|
| A | Quá trình phân lớp |
| B | Quá trình học |
| C | Quá trình dự đoán |
| D | Cây quyết định |

Đáp án đúng: B

Câu hỏi số: 270

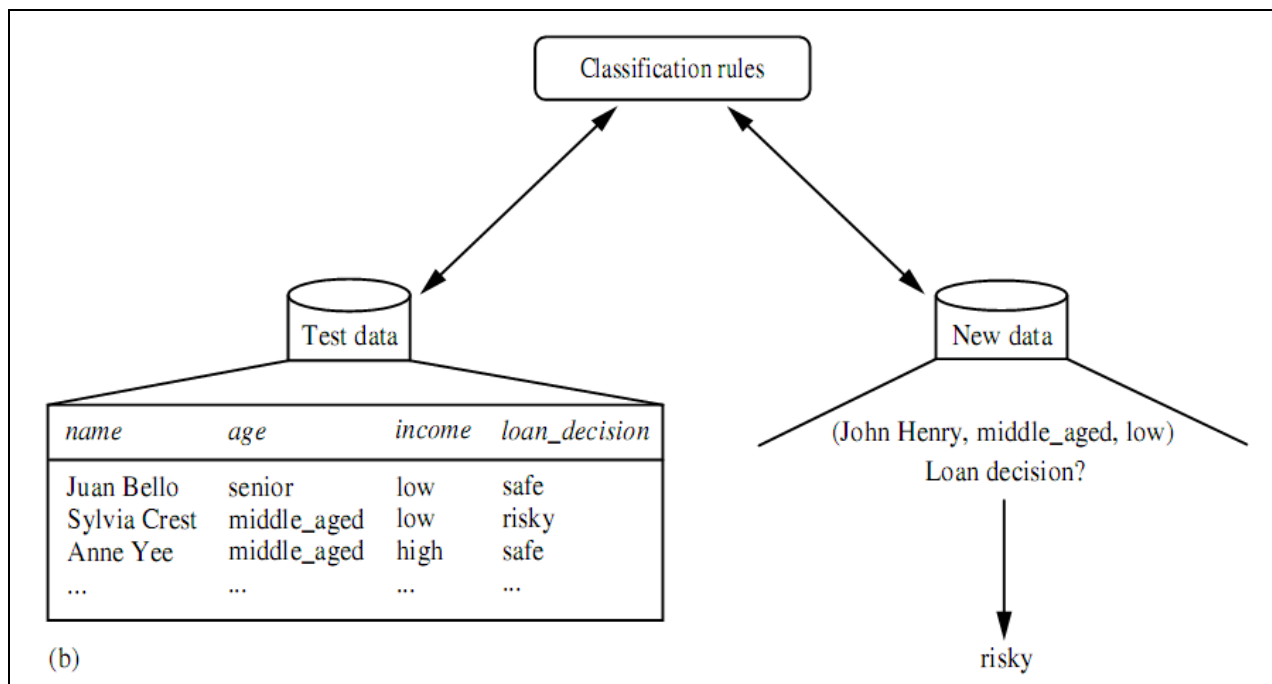
Chương: Tổng hợp

Kỹ năng: Tổng hợp

Mức độ: Dễ

Phân nội dung câu hỏi:

Sơ đồ sau thể hiện quá trình gì?



Các đáp án:

| | |
|---|--------------------|
| A | Quá trình phân lớp |
| B | Quá trình học |
| C | Quá trình dự đoán |
| D | Cây quyết định |

Đáp án đúng: A

Câu hỏi số: 271

Chương: Tổng hợp

Kỹ năng: Tổng hợp

Mức độ: Dễ

Phân nội dung câu hỏi:

| |
|-----------------------------------------------------------------------------|
| Đầu ra trong thuật toán sinh cây quyết định (Generate decision tree) là gì? |
|-----------------------------------------------------------------------------|

Các đáp án:

| | |
|---|-------------------------------------------------------------------------|
| A | Bộ dữ liệu D, trong đó chứa các dữ liệu huấn luyện và các nhãn phân lớp |
| B | Một cây quyết định |
| C | Danh sách các thuộc tính (Attribute list) |
| D | Thủ tục để xác định tiêu chí phân chia các bộ dữ liệu. |

Đáp án đúng: B

Câu hỏi số: 272

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Khó

Phân nội dung câu hỏi:

| |
|------------------------------------------------------------------------------------------------------------------|
| Trong môi trường Explorer của WEKA, để khám phá các luật kết hợp từ dữ liệu, tính năng nào sau đây sẽ được chọn? |
|------------------------------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|-----------|
| A | Visualize |
| B | Classify |

| | |
|---|-----------|
| C | Cluster |
| D | Associate |

Đáp án đúng: D

Câu hỏi số: 273

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Khó

Phân nội dung câu hỏi:

Trong môi trường Explorer của WEKA, để xác định và lựa chọn các thuộc tính liên quan nhất của dữ liệu, tính năng nào sau đây sẽ được chọn?

Các đáp án:

| | |
|---|-------------------|
| A | Visualize |
| B | Classify |
| C | Select attributes |
| D | Associate |

Đáp án đúng: C

Câu hỏi số: 274

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Khó

Phân nội dung câu hỏi:

Trong môi trường Explorer của WEKA, để xem biểu đồ tương tác 2 chiều đối với dữ liệu, tính năng nào sau đây sẽ được chọn?

Các đáp án:

| | |
|---|-------------------|
| A | Visualize |
| B | Classify |
| C | Select attributes |
| D | Associate |

Đáp án đúng: D

Câu hỏi số: 275

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Trung bình

Phân nội dung câu hỏi:

Các công cụ tiền xử lý dữ liệu của WEKA được gọi là filters gồm có?

Các đáp án:

| | |
|---|---------------------------------------------------------------------|
| A | Rời rạc hóa (Discretization) và Chuẩn hóa (Normalization). |
| B | Lấy mẫu (Re-sampling) và Lựa chọn thuộc tính (Attribute selection). |
| C | Chuyển đổi (Transforming) và kết hợp (Combining) các thuộc tính. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 276

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Khó

Phần nội dung câu hỏi:

Dữ liệu tiền xử lý trong WEKA Explorer, nhận được từ những nguồn nào sau đây?

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------|
| A | Dữ liệu có thể được nhập vào (imported) từ một tập tin có khuôn dạng ARFF, CSV. |
| B | Dữ liệu có thể được đọc vào từ một địa chỉ URL. |
| C | Dữ liệu có thể được đọc vào từ một cơ sở dữ liệu thông qua JDBC. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 277

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Khó

Phần nội dung câu hỏi:

Các kỹ thuật phân lớp được hỗ trợ bởi WEKA Explorer gồm có những gì?

Các đáp án:

| | |
|---|--------------------------------------------------------|
| A | Naïve Bayes classifier và Bayesian networks. |
| B | Decision trees và Neural networks. |
| C | Instance-based classifiers và Support vector machines. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 278

Chương: Giới thiệu về WEKA

Kỹ năng: Hiểu

Mức độ: Khó

Phần nội dung câu hỏi:

Các bộ phân lớp (Classifiers) của WEKA tương ứng với các mô hình dự đoán các đại lượng kiểu định danh (phân lớp) hoặc các đại lượng kiểu số (hồi qui/dự đoán). Việc lựa chọn các tùy chọn cho việc kiểm tra (test options) gồm có?

Các đáp án:

| | |
|---|---------------------------------------|
| A | Use training set |
| B | Supplied test set |
| C | Cross-validation và Percentage split. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 279

Chương: Giới thiệu về WEKA

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong các bộ phân lớp của WEKA Explorer, Use training set là gì?

Các đáp án:

| | |
|---|-----------------------------------------------------------------------------|
| A | Bộ phân loại học được sẽ được đánh giá trên tập học. |
| B | Tập dữ liệu khác với tập học được sử dụng cho việc đánh giá. |
| C | Tập dữ liệu sẽ được chia đều thành k tập (folds) có kích thước xấp xỉ nhau. |
| D | Chỉ định tỷ lệ phân chia dữ liệu đối với việc đánh giá. |

Đáp án đúng: C

Câu hỏi số: 280

Chương: Giới thiệu về WEKA

Kỹ năng: Hiểu

Mức độ: Trung bình

Phân nội dung câu hỏi:

| |
|-------------------------------------------------------------------|
| Trong các bộ phân lớp của WEKA Explorer, Supplied test set là gì? |
|-------------------------------------------------------------------|

Các đáp án:

| | |
|---|-----------------------------------------------------------------------------|
| A | Bộ phân loại học được sẽ được đánh giá trên tập học. |
| B | Tập dữ liệu khác với tập học được sử dụng cho việc đánh giá. |
| C | Tập dữ liệu sẽ được chia đều thành k tập (folds) có kích thước xấp xỉ nhau. |
| D | Chỉ định tỷ lệ phân chia dữ liệu đối với việc đánh giá. |

Đáp án đúng: B

Câu hỏi số: 281

Chương: Giới thiệu về WEKA

Kỹ năng: Hiểu

Mức độ: Trung bình

Phân nội dung câu hỏi:

| |
|------------------------------------------------------------------|
| Trong các bộ phân lớp của WEKA Explorer, Cross-validation là gì? |
|------------------------------------------------------------------|

Các đáp án:

| | |
|---|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| A | Bộ phân loại học được sẽ được đánh giá trên tập học. |
| B | Tập dữ liệu khác với tập học được sử dụng cho việc đánh giá. |
| C | Tập dữ liệu sẽ được chia đều thành k tập (folds) có kích thước xấp xỉ nhau, và bộ phân loại học được sẽ được đánh giá bởi phương pháp cross-validation. |
| D | Chỉ định tỷ lệ phân chia dữ liệu đối với việc đánh giá. |

Đáp án đúng: C

Câu hỏi số: 282

Chương: Giới thiệu về WEKA

Kỹ năng: Hiểu

Mức độ: Trung bình

Phân nội dung câu hỏi:

| |
|---------------------------------------------------------------------------------------------|
| Trong các bộ phân lớp của WEKA Explorer, phát biểu nào sau đây là đúng về Percentage split? |
|---------------------------------------------------------------------------------------------|

Các đáp án:

| | |
|---|-----------------------------------------------------------------------------|
| A | Bộ phân loại học được sẽ được đánh giá trên tập học. |
| B | Tập dữ liệu khác với tập học được sử dụng cho việc đánh giá. |
| C | Tập dữ liệu sẽ được chia đều thành k tập (folds) có kích thước xấp xỉ nhau. |
| D | Chỉ định tỷ lệ phân chia dữ liệu đối với việc đánh giá. |

Đáp án đúng: D

Câu hỏi số: 283

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Khó

Phần nội dung câu hỏi:

Classifier output trong WEKA Explorer hiển thị các thông tin quan trọng nào sau đây?

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------------------------------------|
| A | Run information: các tùy chọn đối với mô hình học, tên của tập dữ liệu, số lượng các ví dụ, các thuộc tính. |
| B | Classifier mode: biểu diễn của bộ phân lớp học được. |
| C | Predictions on test data: thông tin chi tiết về các dự đoán của bộ phân lớp đối với tập kiểm tra. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 284

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Khó

Phần nội dung câu hỏi:

Classifier output trong WEKA Explorer hiển thị các thông tin quan trọng nào?

Các đáp án:

| | |
|---|------------------------------------------------------------------------------------------------------------------------------|
| A | Summary: các thống kê về mức độ chính xác của bộ phân lớp |
| B | Detailed Accuracy By Class: thông tin chi tiết về mức độ chính xác của bộ phân lớp đối với mỗi lớp. |
| C | Confusion Matrix: các thành phần của ma trận này thể hiện số lượng các ví dụ kiểm tra được phân lớp đúng và bị phân lớp sai. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 285

Chương: Giới thiệu về WEKA

Kỹ năng: Hiểu

Mức độ: Trung bình

Phần nội dung câu hỏi:

Result list trong WEKA Explorer cung cấp một số chức năng hữu ích như là?

Các đáp án:

| | |
|---|--------------------------------------------------------------------------------------------------------------|
| A | Save model: lưu lại mô hình tương ứng với bộ phân lớp học được vào trong một tập tin nhị phân. |
| B | Load model: đọc lại một mô hình đã được học trước đó từ một tập tin nhị phân. |
| C | Re-evaluate model on current test set: đánh giá một mô hình học được trước đó đối với tập kiểm tra hiện tại. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 286

Chương: Giới thiệu về WEKA

Kỹ năng: Biết

Mức độ: Khó

Phản nội dung câu hỏi:

Phát biểu nào sau đây về các bộ phân cụm (Cluster builders) của WEKA Explorer là đúng?

Các đáp án:

| | |
|---|--------------------------------------------------------------------------------------------------------------------------------|
| A | Các bộ phân cụm (Cluster builders) của WEKA tương ứng với các mô hình tìm các nhóm của các ví dụ tương tự đối với tập dữ liệu. |
| B | EM, K-Means là các kỹ thuật phân cụm được hỗ trợ bởi WEKA. |
| C | Các bộ phân cụm có thể được hiển thị kết quả và so sánh với các cụm (lớp) thực tế. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 287

Chương: Giới thiệu về WEKA

Kỹ năng: Áp dụng

Mức độ: Khó

Phản nội dung câu hỏi:

Lựa chọn chế độ phân cụm (Cluster mode) trong WEKA Explorer gồm có?

Các đáp án:

| | |
|---|-------------------------------------------------------------------------------|
| A | Use training set: các cụm học được sẽ được kiểm tra đối với tập học. |
| B | Supplied test set: sử dụng một tập dữ liệu khác để kiểm tra các cụm học được. |
| C | Percentage split và Classes to cluster evaluation. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 288

Chương: Giới thiệu về WEKA

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phản nội dung câu hỏi:

Làm việc với các bộ phân cụm trong WEKA Explorer: khi muốn lưu lại các bộ phân lớp trong bộ nhớ, để có thể hiển thị sau đó, chức năng nào sau đây được lựa chọn?

Các đáp án:

| | |
|---|----------------------------------|
| A | Store cluster for visualization. |
| B | Ignore attributes. |
| C | Percentage split. |
| D | Classes to cluster evaluation. |

Đáp án đúng: A

Câu hỏi số: 289

Chương: Giới thiệu về WEKA

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phản nội dung câu hỏi:

Làm việc với các bộ phân cụm trong WEKA Explorer: khi muốn lựa chọn các thuộc tính sẽ không tham gia vào quá trình học các cụm, chức năng nào sau đây được lựa chọn?

Các đáp án:

| | |
|---|----------------------------------|
| A | Store cluster for visualization. |
| B | Ignore attributes. |
| C | Percentage split. |
| D | Classes to cluster evaluation. |

Đáp án đúng: B

Câu hỏi số: 290

Chương: Giới thiệu về WEKA

Kỹ năng: Áp dụng

Mức độ: Khó

Phân nội dung câu hỏi:

Khi lựa chọn chế độ phân cụm (cluster mode), ở mode nào các cụm học được sẽ được kiểm tra đối với tập học?

Các đáp án:

| | |
|---|-------------------------------|
| A | Use training set |
| B | Supplied test set |
| C | Percentage split |
| D | Classes to cluster evaluation |

Đáp án đúng: A

Câu hỏi số: 291

Chương: Giới thiệu về WEKA

Kỹ năng: Áp dụng

Mức độ: Khó

Phân nội dung câu hỏi:

Khi lựa chọn chế độ phân cụm (cluster mode), ở mode nào sử dụng một tập dữ liệu khác để kiểm tra các cụm học được?

Các đáp án:

| | |
|---|-------------------------------|
| A | Use training set |
| B | Supplied test set |
| C | Percentage split |
| D | Classes to cluster evaluation |

Đáp án đúng: B

Câu hỏi số: 292

Chương: Giới thiệu về WEKA

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phân nội dung câu hỏi:

Khi lựa chọn chế độ phân cụm (cluster mode), ở mode nào chỉ định tỉ lệ phân chia tập dữ liệu ban đầu cho việc xây dựng tập kiểm tra?

Các đáp án:

| | |
|---|-------------------------------|
| A | Use training set |
| B | Supplied test set |
| C | Percentage split |
| D | Classes to cluster evaluation |

Đáp án đúng: C

Câu hỏi số: 293

Chương: Giới thiệu về WEKA

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

Khi lựa chọn chế độ phân cụm (cluster mode), ở mode nào so sánh độ chính xác của các cụm học được đối với các lớp được chỉ định?

Các đáp án:

| | |
|---|-------------------------------|
| A | Use training set |
| B | Supplied test set |
| C | Percentage split |
| D | Classes to cluster evaluation |

Đáp án đúng: C

Câu hỏi số: 294

Chương: Giới thiệu về WEKA

Kỹ năng: Phân tích

Mức độ: Trung bình

Phần nội dung câu hỏi:

Khi lựa chọn một mô hình (giải thuật) phát hiện luật kết hợp, trong WEKA Explorer. Associator output hiển thị các thông tin quan trọng gì?

Các đáp án:

| | |
|---|--------------------------------------------------------------------------------|
| A | Độ hỗ trợ tối thiểu |
| B | Độ tin cậy tối thiểu |
| C | Kích thước của các tập mục thường xuyên, và liệt kê các luật kết hợp tìm được. |
| D | Cả 3 đáp án trên đều đúng. |

Đáp án đúng: D

Câu hỏi số: 295

Chương: Giới thiệu về WEKA

Kỹ năng: Phân tích

Mức độ: Khó

Phần nội dung câu hỏi:

Việc lựa chọn thuộc tính trong WEKA Explorer nhằm mục đích gì?

Các đáp án:

| | |
|---|------------------------------------------------------|
| A | Để lựa chọn một mô hình phát hiện liên kết. |
| B | Để lựa chọn chế độ phân cụm. |
| C | Để xác định những thuộc tính nào là quan trọng nhất. |
| D | Để cung cấp một số chức năng hữu ích. |

Đáp án đúng: C

Câu hỏi số: 296

Chương: Giới thiệu về WEKA

Kỹ năng: Áp dụng

Mức độ: Trung bình

Phần nội dung câu hỏi:

Trong WEKA, một phương pháp lựa chọn thuộc tính (attribute selection) bao gồm mấy phần?

Các đáp án:

| | |
|---|---|
| A | 1 |
| B | 2 |
| C | 3 |
| D | 4 |

Đáp án đúng: B

Câu hỏi số: 297

Chương: Giới thiệu về WEKA

Kỹ năng: Áp dụng

Mức độ: Khó

Phân nội dung câu hỏi:

Trong WEKA, để xác định một phương pháp đánh giá mức độ phù hợp của các thuộc tính như correlation-based, wrapper, information gain, chi-squared,... chức năng nào sẽ được lựa chọn?

Các đáp án:

| | |
|---|---------------------|
| A | Supplied test set |
| B | Use training set |
| C | Search Method |
| D | Attribute Evaluator |

Đáp án đúng: D

Câu hỏi số: 298

Chương: Giới thiệu về WEKA

Kỹ năng: Áp dụng

Mức độ: Khó

Phân nội dung câu hỏi:

Trong WEKA, để xác định một phương pháp (thứ tự) xét các thuộc tính như best-first, random, exhaustive, ranking,... tính năng nào được lựa chọn?

Các đáp án:

| | |
|---|---------------------|
| A | Supplied test set |
| B | Use training set |
| C | Search Method |
| D | Attribute Evaluator |

Đáp án đúng: C