УТВЕРЖДАЮ

Зав. кафедрой СП ЮУрГУ

\_\_\_\_\_\_\_\_\_Л.Б. Соколинский

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_2022 г.

Фонд оценочных средств ООП «Искусственный интеллект и инженерия данных» по направлению 09.04.04 – Программная инженерия

Дисциплина «Основы интеллектуального анализа данных»

| **№ КМ** | **Вид КМ** | **Наименование КМ** | **Оценочные средства** |
| --- | --- | --- | --- |
|  | Текущий контроль | Тест 1  "Введение в дисциплину" | Пример вопросов теста:  1) Установите соответствие между базовыми задачами интеллектуального анализа данных и приведенными задачами реальной предметной области.   |  |  |  | | --- | --- | --- | | Определение списка клиентов банка, имеющих задолженности по выплатам кредита |  | Задача НЕ из области интеллектуального анализа данных | | Определение смысловых групп клиентов банка, взявших кредит, на основе персональных данных этих клиентов |  | Задача кластеризации | | Определение пакета услуг, которые часто выбираются клиентами банка совместно с взятием кредита |  | Задача классификации | | Определение возможности выдачи кредита клиенту банка на основе персональных данных этого клиента |  | Задача поиска ассоциативных правил |   2) Пронумеруйте этапы цикла аналитической обработки данных (извлечение знаний из данных), начиная с 1.   |  |  | | --- | --- | |  | Очистка данных | |  | Интерпретация результатов | |  | Интеграция источников данных | |  | Загрузка данных в хранилище | |  | Выполнение интеллектуального и (или) оперативного анализа данных |   3) Укажите ТРИ наиболее важные характеристики феномена Big Data (Большие данные):   |  |  | | --- | --- | |  | Velocity (скорость прироста) | |  | Verifyability (верифицируемость) | |  | Vanity (тщеславие) | |  | Variety (разнообразие) | |  | Visualization (наглядное представления результатов) | |  | Value (ценность результатов) | |  | Viability (жизнеспособность) | |  | Veracity (достоверность) | |  | Volume (размер) | |  | Validity (обоснованность) | |
|  | Текущий контроль | Тест 2  "Поиск шаблонов" | Пример вопросов теста:  1) Укажите верное определение достоверности шаблона A→B, где A≠∅,B≠∅,A∩B=∅A≠∅,B≠∅,A∩B=∅.    2) Укажите верное утверждение о связи между количеством частых наборов и порогом minsup (минимальной поддержкой).   |  |  | | --- | --- | |  | Увеличение минимальной поддержки не ведет к изменению количества частых наборов | |  | Увеличение минимальной поддержки ведет к уменьшению количества частых наборов | |  | Увеличение минимальной поддержки ведет к случайному изменению количества частых наборов | |  | Увеличение минимальной поддержки ведет к увеличению количества частых наборов |   3) Укажите верное определение поддержки шаблона A→B, где A≠∅,B≠∅,A∩B=∅A≠∅,B≠∅,A∩B=∅. |
|  | Текущий контроль | Тест 3  "Классификация" | Пример вопросов теста:  1) Укажите верное определение обучающей выборки для задачи классификации.   |  |  | | --- | --- | |  | Пересечение множеств, используемых для построения и проверки модели классификации | |  | Множество классифицированных объектов, классификация которых должна быть выполнена на основе построенной модели для ее проверки | |  | Множество не классифицированных кортежей, классификация которых должна быть выполнена на основе построенной модели. | |  | Множество классифицированных объектов, используемых для построения модели классификации |   2) Укажите верную последовательность выполнения этапов процесса классификации.   |  |  | | --- | --- | |  | Построение модели на основе тестовой выборки → Оценка точности модели на основе обучающей выборки → Классификация ранее неизвестных данных | |  | Построение модели на основе обучающей выборки → Оценка точности модели на основе тестовой выборки → Классификация ранее неизвестных данных | |  | Оценка точности модели на основе обучающей выборки → Оценка точности модели на основе тестовой выборки → Классификация ранее неизвестных данных | |  | Оценка точности модели на основе тестовой выборки → Построение модели на основе обучающей выборки → Классификация ранее неизвестных данных | |  | Оценка точности модели на основе обучающей выборки → Построение модели на основе тестовой выборки → Классификация ранее неизвестных данных |   3) Укажите две задачи реальной предметной области, которые являются задачами классификации.   |  |  | | --- | --- | |  | Определением смысловых групп писем электронной почты на основе данных о ключевых словах этих писем. | |  | Определение списка корреспондентов, отправивших наибольшее количество электронных писем. | |  | Определение категории письма электронной почты: "спам" или "обычная почта" – на основе данных о ключевых словах этого письма. | |  | Определение смысловых групп клиентов банка, взявших кредит, на основе персональных данных этих клиентов. | |  | Определение возможности выдачи кредита клиенту банка на основе персональных данных этого клиента. | |  | Определение адресов электронной почты, которые часто фигурируют совместно в списке адресатов писем. | |  | Определение списка клиентов банка, имеющих задолженности по выплатам кредита. | |  | Определение пакета услуг, которые часто выбираются клиентами банка совместно с взятием кредита. | |
|  | Текущий контроль | Тест 4  "Кластеризация" | Пример вопросов теста:  1) Укажите верное определение задачи кластеризации.   |  |  | | --- | --- | |  | Разделение объектов заданного множества на два кластера: те, что могут быть использованы для принятия стратегических решений, и остальные | |  | Определение кластеров, к которым принадлежат объекты заданного множества, по характеристикам этих объектов. Семантика кластеров, к которым может быть отнесен объект, заранее не известна | |  | Нахождение часто встречающихся зависимостей между кластерами объектов | |  | Определение кластеров, к которым принадлежат объекты заданного множества, по характеристикам этих объектов. Семантика кластеров, к которым может быть отнесен объект, заранее известна |   2) Укажите основную идею разделительных алгоритмов кластеризации   |  |  | | --- | --- | |  | Предполагается, что каждый исходный объект образует отдельный кластер, и затем выполняется слияние близких друг к другу объектов или кластеров до тех пор, пока не будет получен единственный кластер или не будет выполнено условие завершения слияния | |  | Добавление объектов в кластер до тех пор, пока количество соседних объектов не превысит некоторого заданного порога; при этом в окрестности каждого объекта кластера должно находиться некоторое минимальное количество других объектов | |  | Кластеризация выполняется в два этапа:  1) разбиение исходного множества объектов на кластеры (в каждом кластере имеется, по крайней мере, один объект и каждый объект принадлежит в точности одному кластеру);  2) итеративное перемещение объектов между кластерами с целью улучшить начальное разбиение (чтобы объекты из одного кластера были более "близкими", а из разных кластеров – более "далекими"). | |  | Предполагается, что все исходные объекты входят в один кластер, и затем итеративно выполняется его разбиение на менее мощные кластеры до тех пор, пока не будут получены кластеры-синглтоны или не будет выполнено условие завершения разбиения |   3) Укажите основную идею плотностных алгоритмов кластеризации   |  |  | | --- | --- | |  | Предполагается, что все исходные объекты входят в один кластер, и затем итеративно выполняется его разбиение на менее мощные кластеры до тех пор, пока не будут получены кластеры-синглтоны или не будет выполнено условие завершения разбиения | |  | Предполагается, что каждый исходный объект образует отдельный кластер, и затем выполняется слияние близких друг к другу объектов или кластеров до тех пор, пока не будет получен единственный кластер или не будет выполнено условие завершения слияния | |  | Добавление объектов в кластер до тех пор, пока количество соседних объектов не превысит некоторого заданного порога; при этом в окрестности каждого объекта кластера должно находиться некоторое минимальное количество других объектов | |  | Кластеризация выполняется в два этапа:  1) разбиение исходного множества объектов на кластеры (в каждом кластере имеется, по крайней мере, один объект и каждый объект принадлежит в точности одному кластеру);  2) итеративное перемещение объектов между кластерами с целью улучшить начальное разбиение (чтобы объекты из одного кластера были более "близкими", а из разных кластеров – более "далекими"). | |
|  | Текущий контроль | Тест 5  "Поиск аномалий" | Пример вопросов теста:  1) Укажите верную легенду для следующего графика, отображающего модельные данные    2) Укажите утверждение, верно отражающее правило трех сигм для поиска аномалий в одномерных данных, имеющих нормальное распределение.   |  |  | | --- | --- | |  | Аномалией является точка, для которой отношение модуля разности этой точки и среднего значения выборки к среднеквадратическому отклонению не больше трех | |  | Аномалией является точка, для которой модуль разности этой точки и утроенного среднего значения выборки не больше наперед заданного аналитиком параметра | |  | Аномалией является точка, для которой отношение модуля разности этой точки и среднего значения выборки к среднеквадратическому отклонению не меньше трех | |  | Аномалией является точка, для которой модуль разности этой точки и утроенного среднего значения выборки не меньше наперед заданного аналитиком параметра |   3) Укажите алгоритм кластеризации, который в силу чувствительности к шумам и выбросам в данных даст худшие (среди перечисленных алгоритмов) результаты при его применении для поиска аномалий.   |  |  | | --- | --- | |  | иерархическая кластеризация | |  | k-means | |  | DBSCAN | |  | Fuzzy c-Means | |
|  | Текущий контроль | Поиск шаблонов. Поиск частых наборов | Вопросы для подготовки к устному опросу:  1. Дайте определения следующих понятий: база транзакций, поддержка, частый набор.  2. Объясните принцип антимонотонности поддержки.  3. Объясните работу одного из следующих алгоритмов поиска частых наборов: Apriori, FP Growth, ECLAT.  4. Объясните идею и схему использования фрагментации базы транзакций для поиска частых наборов.  5. Объясните идею и схему использования сэмплинга базы транзакций для поиска частых наборов. |
|  | Текущий контроль | Поиск шаблонов. Поиск ассоциативных правил | Вопросы для подготовки к устному опросу:  1. Дайте определения следующих понятий: шаблон (ассоциативное правило), достоверность, устойчивый шаблон.  2. Объясните алгоритм поиска устойчивых шаблонов с помощью поиска частых наборов.  3. Приведите пример устойчивого, но практически бесполезного шаблона.  4. Дайте определение меры lift полезности шаблонов.  5. Дайте определения понятий максимально частого и замкнутого набора, иерархии наборов. |
|  | Текущий контроль | Классификация. Байесовская классификация | Вопросы для подготовки к устному опросу:  1. Объясните постановку задачи классификации.  2. Объясните типовой процесс классификации и его этапы: обучение модели, оценка модели, применение модели.  3. Сформулируйте теорему Байеса.  4. Объясните, почему на практике, как правило, Байесовская классификация дает невысокую точность классификации.  5. Сформулируйте проблему нулевой вероятности в Байесовской классификации и ее решение. |
|  | Текущий контроль | Классификация. Деревья решений | Вопросы для подготовки к устному опросу:  1. Объясните принцип построения дерева решений.  2. Дайте определения критерия выбора атрибута разбиения Information Gain.  3. Дайте определение критерия выбора атрибута разбиения Gain Ratio.  4. Дайте определение критерия выбора атрибута разбиения Gini Index.  5. Дайте определения показателей качества классификации: аккуратность (accuracy), точность (precision), полнота (recall), F-мера. |
|  | Текущий контроль | Классификация. Бэггинг | Вопросы для подготовки к устному опросу:  1. Объясните цель и идею ансамблевой классификации.  2. Объясните работу метода бэггинга.  3. Объясните, допустимо ли вхождение в ансамбль, выполняющий бэггинг, разнородных классификаторов.  4. Назовите преимущества и недостатки метода бэггинга.  5. Объясните, почему бэггинг предполагает примерную вероятность 0.632 включения элемента исходной обучающей выборки в выборку участника ансамбля. |
|  | Текущий контроль | Классификация. Случайный лес | Вопросы для подготовки к устному опросу:  1. Объясните цель и идею ансамблевой классификации.  2. Объясните работу метода случайного леса.  3. Назовите преимущества и недостатки метода случайного леса.  4. Объясните работу одной из разновидностей метода случайного леса, Forest-RI.  5. Объясните работу одной из разновидностей метода случайного леса, Forest-RC. |
|  | Текущий контроль | Классификация. Бустинг | Вопросы для подготовки к устному опросу:  1. Объясните цель и идею ансамблевой классификации.  2. Объясните работу метода бустинга (на примере алгоритма AdaBoost).  3. Назовите преимущества и недостатки метода бустинга.  4. Укажите ошибку обучения ансамбля классификаторов в методе бустинга.  5. Объясните, как убывает ошибка обучения ансамбля в бустинге при увеличении количества классификаторов в ансамбле. |
|  | Текущий контроль | Кластеризация. Разделительная кластеризация | Вопросы для подготовки к устному опросу:  1. Объясните идею разделительной кластеризации и работу алгоритма k Means.  2. Напишите формулу меры для выявления кластеров в k Means (Sum of Squared Errors)  3. Назовите преимущества и недостатки алгоритма k Means.  4. Объясните один из способов (на выбор) подбора начальных центроидов в алгоритме k Means.  5. Объясните работу алгоритма k Medoids. |
|  | Текущий контроль | Кластеризация. Плотностная кластеризация | Вопросы для подготовки к устному опросу:  1. Объясните идею плотностной кластеризации и работу алгоритма DBSCAN.  2. Дайте определения основных понятий, используемых в алгоритме DBSCAN: окрестность точки, корневая точка, непосредственная достижимость, достижимость.  3. Назовите преимущества и недостатки алгоритма DBSCAN.  4. Объясните, почему алгоритм DBSCAN является нечувствительным выбросам и шумам в исходных данных.  5. Объясните, каким можно подбирать параметры и алгоритма DBSCAN. |
|  | Текущий контроль | Кластеризация. Иерархическая кластеризация | Вопросы для подготовки к устному опросу:  1. Объясните идею иерархической кластеризации и работу алгоритма агломеративной кластеризации.  2. Объясните идею иерархической кластеризации и работу алгоритма дивизимной кластеризации.  3. Объясните способ построения дендрограмм.  4. Назовите преимущества и недостатки иерархической кластеризации.  5. Дайте определения следующих мер схожести, используемых в иерархической кластеризации: Single linkage, Complete linkage, Average linkage. |
|  | Текущий контроль | Кластеризация. Качество кластеризации | Вопросы для подготовки к устному опросу:  1. Объясните работу метода локтя определения оптимального количества кластеров.  2. Объясните работу метода кросс-валидации определения оптимального количества кластеров.  3. Дайте определение силуэтного коэффициента и объясните его применение для определения оптимального количества кластеров.  4. Объясните способ оценки неслучайности кластеризуемых данных на основе числа Хопкинса.  5. Объясните способ оценки качества кластеризации на основе предварительной классификации. |
|  | Текущий контроль | Поиск аномалий. Поиск точечных аномалий | Вопросы для подготовки к устному опросу:  1. Дайте определение понятия аномалии и разновидностей аномалий (точечная, коллективная, контекстная).  2. Объясните работу метода максимального правдоподобия для поиска аномалий.  3. Объясните работу метода построения гистограмм для поиска аномалий.  4. Объясните работу метода z-значимости для поиска аномалий.  5. Объясните работу правила трех сигм для поиска аномалий. |
|  | Текущий контроль | Поиск аномалий. Поиск коллективных аномалий | Вопросы для подготовки к устному опросу:  1. Дайте определение понятия аномалии и разновидностей аномалий (точечная, коллективная, контекстная).  2. Объясните работу метода вложенных циклов для поиска аномалий.  3. Объясните работу метода решеток для поиска аномалий.  4. Дайте определение понятия диссонанса временного ряда.  5. Объясните принцип поиска аномалий на основе кластеризации данных. |
|  | Промежуточный контроль | Итоговый тест | Примеры вопросов итогового теста:  1) Укажите термин, который соответствует следующему определению феномена современного информационного общества:  Наличие разнообразных данных, имеющих большой объем и высокую скорость прироста, и требующих специализированных аппаратных, программных и алгоритмических средств для получения на основе этих данных полезных знаний.   |  |  | | --- | --- | |  | Сверхбольшие базы данных (Very Large Databases) | |  | Большие данные (Big Data) | |  | Большие базы данных (Big Databases) | |  | Большие массивы данных (Massive Datasets) |   2) Укажите временной период возникновения и становления термина Big Data.   |  |  | | --- | --- | |  | 2000-2005 гг. | |  | 2005-2010 гг. | |  | 2010-2015 гг. | |  | 1995-2000 гг. |   3) Укажите задачу, которая относится к сфере Data Mining.   |  |  | | --- | --- | |  | Нахождение разбиения клиентов компании на группы в зависимости величины полученной от них прибыли | |  | Нахождение общей прибыли компании, полученной ей от клиентов | |  | Нахождение ожидаемой прибыли от нового клиента компании на основе данных о существующих клиентах | |  | Нахождение разбиения клиентов компании на две группы: мужчины и женщины | |  | Нахождение списка клиентов компании, фактическая прибыль от которых меньше ожидаемой |   4) Укажите верное определение термина Data Cleaning (очистка данных).   |  |  | | --- | --- | |  | Перемещение данных, подвергаемых интеллектуальному анализу, из источников данных в хранилище данных. | |  | Выявление и исправление ошибок, несоответствий данных с целью улучшения их качества. | |  | Отбор данных предметной области, подвергаемых интеллектуальному анализу. | |  | Отбор результатов интеллектуального анализа данных, полезных в предметной области. |   5) Укажите верное определение задачи шаблонов и ассоциативных правил.   |  |  | | --- | --- | |  | Определение, какие из имеющихся данных могут быть использованы для принятия стратегических решений, а какие – нет. | |  | Определение классов, к которым принадлежат объекты заданного множества, по характеристикам этих объектов. Множество классов, к которым может быть отнесен объект, заранее известно. | |  | Определение классов, к которым принадлежат объекты заданного множества, по характеристикам этих объектов. Множество классов, к которым может быть отнесен объект, заранее не известно. | |  | Нахождение часто встречающихся зависимостей между объектами. |   6) Укажите верное определение задачи поиска частых наборов (анализ рыночных корзин).   |  |  | | --- | --- | |  | Обнаружение наборов товаров, которые редко приобретаются совместно. | |  | Обнаружение наборов товаров, которые приобретаются совместно с товарами, приобретаемыми наиболее редко. | |  | Обнаружение наборов товаров, которые приобретаются совместно с товарами, приобретаемыми наиболее часто. | |  | Обнаружение наборов товаров, которые приобретаются наиболее редко. | |  | Обнаружение наборов товаров, которые часто приобретаются совместно. | |  | Обнаружение наборов товаров, которые приобретаются наиболее часто. |   7) Среди данных шаблонов, найденных при некоторых заданных пороговых значениях minsup и minconf, укажите те, которые будут иметь практическую пользу.   |  |  | | --- | --- | |  | Шаблон хлеб → авокадо, который имеет высокую поддержку и высокую достоверность | |  | Шаблон хлеб → нож, который имеет высокую поддержку и низкую достоверность | |  | Шаблон хлеб → стиральный порошок, который имеет низкую поддержку и низкую достоверность | |  | Шаблон хлеб → масло, который имеет высокую поддержку и высокую достоверность | |  | Шаблон водка → икра, который имеет низкую поддержку и высокую достоверность |   8) Укажите все замкнутые 1-наборы, если minsup=2.  /var/folders/71/j5hjxk7x3fg2h9tl8qp12bdr0000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/wGL4P7WqPQxyQAAAABJRU5ErkJggg==   |  |  | | --- | --- | |  | D | |  | C | |  | A | |  | E | |  | B |   9) Укажите способ генерации множества кандидатов в частые 2-наборы в алгоритме Apriori.    10) Укажите все замкнутые частые наборы, если minsup=2.  /var/folders/71/j5hjxk7x3fg2h9tl8qp12bdr0000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/8H4eB78BLu3HsAAAAASUVORK5CYII=   |  |  | | --- | --- | |  | E | |  | CE | |  | CD | |  | C | |  | CDE | |  | DE | |  | D |   11) Укажите верное определение термина "подгонка" (overfitting) для задачи классификации.   |  |  | | --- | --- | |  | Множество не классифицированных кортежей, классификация которых должна быть выполнена на основе построенной модели. | |  | Множество классифицированных объектов, используемых для построения модели классификации. | |  | Множество классифицированных объектов, классификация которых должна быть выполнена на основе построенной модели для ее проверки. | |  | Пересечение множеств, используемых для построения и проверки модели классификации. |   12) Укажите основную идею алгоритма классификации С4.5.   |  |  | | --- | --- | |  | Для классификации множества объектов используется значение меток классов у заданного количества кортежей обучающей выборки, которые в смысле заданной метрики находятся близко от кортежей тестовой выборки. | |  | Для классификации множества объектов выбирается атрибут, разбиение заданного множества по которому дает наименьшее значение индекса Джини (Gini Index). | |  | Для классификации множества объектов выбирается атрибут, разбиение заданного множества по которому дает наибольший прирост информации (Information Gain). | |  | Для классификации множества объектов выбирается атрибут, разбиение заданного множества по которому дает наибольшее соотношение прироста информации и информации, необходимой для разбиения (Gain Ratio). |   13) Вычислите точность (Precision) бинарного классификатора (классы: + и -), имеющего следующую матрицу ошибок (строки – прогнозные данные, столбцы – реальные данные):  /var/folders/71/j5hjxk7x3fg2h9tl8qp12bdr0000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/gfzroxlRy4WYwAAAABJRU5ErkJggg==  14) Вычислите полноту (Recall) бинарного классификатора (классы: + и -), имеющего следующую матрицу ошибок (строки – прогнозные данные, столбцы – реальные данные):  /var/folders/71/j5hjxk7x3fg2h9tl8qp12bdr0000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/mlfD9Tamab1LnTm495F8u1rPhpbj61i2toAX5MACUSbwIHvtNGeVc4dCZDADgGaln1AAh1G4P8BzNjed7NsjkgAAAAASUVORK5CYII=  15) Укажите основное назначение ансамблевой классификации.   |  |  | | --- | --- | |  | Увеличение быстродействия классификации | |  | Увеличение точности классификации | |  | Уменьшение объема обучающей выборки для классификации | |  | Уменьшение объема тестовой выборки для классификации |   16) Укажите основную идею агломеративных алгоритмов кластеризации.   |  |  | | --- | --- | |  | Добавление объектов в кластер до тех пор, пока количество соседних объектов не превысит некоторого заданного порога; при этом в окрестности каждого объекта кластера должно находиться некоторое минимальное количество других объектов. | |  | Кластеризация выполняется в два этапа:  1) разбиение исходного множества объектов на кластеры (в каждом кластере имеется, по крайней мере, один объект и каждый объект принадлежит в точности одному кластеру);  2) итеративное перемещение объектов между кластерами с целью улучшить начальное разбиение (чтобы объекты из одного кластера были более "близкими", а из разных кластеров – более "далекими"). | |  | Предполагается, что каждый исходный объект образует отдельный кластер, и затем выполняется слияние близких друг к другу объектов или кластеров до тех пор, пока не будет получен единственный кластер или не будет выполнено условие завершения слияния. | |  | Предполагается, что все исходные объекты входят в один кластер, и затем итеративно выполняется его разбиение на менее мощные кластеры до тех пор, пока не будут получены кластеры-синглтоны или не будет выполнено условие завершения разбиения. |   17) Укажите вид алгоритмов кластеризации, к которому относится алгоритм k-means.   |  |  | | --- | --- | |  | Агломеративные алгоритмы | |  | Дивизимные алгоритмы | |  | Разделительные алгоритмы | |  | Нечеткие алгоритмы | |  | Плотностные алгоритмы |   18) Укажите верный результат кластеризации объектов множества {4, 5, 8, 9, 10} посредством алгоритма k-means при k=2.   |  |  | | --- | --- | |  | С1={4, 5, 8, 9} C2={10} | |  | С1={4, 5} C2={8, 9, 10} | |  | С1={4, 5, 8} C2={9, 10} | |  | С1={4, 5, 10} C2={8, 9} |   19) Укажите преимущество, которым (в общем случае) обладает алгоритм кластеризации k-medoids по сравнению с алгоритмом k-means.   |  |  | | --- | --- | |  | Большая устойчивость к шумам и выбросам в данных. | |  | Большая скорость работы (более быстрая сходимость). | |  | Лучшее качество кластеризации. | |  | Возможность распараллеливания. |   20) Укажите оптимальное количество кластеров на основе следующего графика зависимости SSE (Sum of Squared Errors) от количества кластеров.  /var/folders/71/j5hjxk7x3fg2h9tl8qp12bdr0000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/x+46mcFRnTkpQAAAABJRU5ErkJggg==   |  |  | | --- | --- | |  | 3 | |  | 5 | |  | 4 | |  | 2 | |  | 8 | |  | 7 | |  | 1 |   21) Укажите статистическую функцию, которая является наименее чувствительной (среди остальных) к аномалиям (выбросам) в данных.   |  |  | | --- | --- | |  | среднее | |  | стандартное отклонение | |  | медиана | |  | мода |   22) Пусть к непустому множеству объектов применен некий алгоритм поиска аномалий. Затем к полученному непустому подмножеству аномальных объектов вновь применен первоначальный алгоритм. Укажите верное утверждение.   |  |  | | --- | --- | |  | Безусловно, все полученные аномальные объекты будут вновь классифицированы как аномалии | |  | Безусловно, ни один из полученных аномальных объектов не будет вновь классифицирован как аномалия | |  | Все полученные аномальные объекты будут вновь классифицированы как аномалии, если алгоритм использует понятие плотности объектов и те же пороговые значения, что на первом шаге |   23) Пусть вероятность распознавания нормального объекта как аномалии 1% и вероятность распознавания аномального объекта как аномалии 99%. Пусть 99% объектов выборки являются нормальными и 1% объектов выборки аномалии. Вычислите в % частоту ложного срабатывания (отношение количества ложных аномалий к количеству объектов, классифицированных как аномалии).  24) Пусть имеется непустой набор точек, имеющих заданное распределение на отрезке [0,1]. Укажите распределение для этих данных, при котором будет иметь смысл понятие выброса как редко наблюдаемой величины?   |  |  | | --- | --- | |  | любое, кроме равномерного | |  | только нормальное | |  | любое, кроме нормального | |  | нормальное или равномерное | |  | только равномерное |   25) Укажите верное определение диссонанса (discord) временного ряда.   |  |  | | --- | --- | |  | Диссонанс – подпоследовательность ряда, расстояние от которой до ее ближайшего соседа максимально. Ближайший сосед подпоследовательности – подпоследовательность ряда, которая не пересекается с данной и имеет минимальное расстояние до нее. | |  | Диссонанс – подпоследовательность ряда, расстояние от которой до ее ближайшего соседа минимально. Ближайший сосед подпоследовательности – подпоследовательность ряда, которая не пересекается с данной и имеет максимальное расстояние до нее. | |  | Диссонанс – подпоследовательность ряда, расстояние от которой до ее ближайшего соседа максимально. Ближайший сосед подпоследовательности – подпоследовательность ряда, которая  имеет минимальное расстояние до нее. | |  | Диссонанс – подпоследовательность ряда, расстояние от которой до ее ближайшего соседа минимально. Ближайший сосед подпоследовательности – подпоследовательность ряда, которая имеет максимальное расстояние до нее. | |  | Диссонанс – подпоследовательность ряда, каждая точка которой является выбросом в соответствии с правилом трех сигм. | |

Паспорт фонда оценочных средств приведен в п. 6.3 РПД.

Разработчик М.Л. Цымблер