



Введение в анализ больших данных.oi(dor_БАК)

- 1 Variety в контексте характеристик Big Data означает ...
- 2 Неверно, что Big Data – это ...
- 3 Отказоустойчивость Big Data имеет место, когда ...
- 4 В числе примеров структурированных данных – ... (укажите 2 варианта ответа)
- 5 Название технологии, которая имитирует человеческое поведение, – искусственный ...
- 6 Компьютер, выделенный из группы персональных компьютеров (или рабочих станций) для выполнения какой-либо сервисной задачи без непосредственного участия человека, – это ...
- 7 Совокупность сведений, зафиксированных на определенном носителе в форме, пригодной для постоянного хранения, передачи и обработки, – это ...
- 8 Целое четырехзначное число, отражающее количество петабайт в зеттабайте, – ...
- 9 Установите соответствие целей инфраструктуры Hadoop и их характеристик:
- 10 Установите правильную последовательность этапов процесса MapReduce подсчета цветов в текстовых файлах:
- 11 Аналитическим отделом компании был получен спам по e-mail с приложенным файлом: Что следует предпринять?
- 12 Реляционные базы данных организованы в виде ...
- 13 SQL – это ...
- 14 Hadoop – это проект ...
- 15 В числе систем управления базами данных (СУБД), которые полностью полагаются на оперативную память при хранении информации, – ... (укажите 2 варианта ответа)





- 16) Английское название фреймворка для вычисления наборов распределенных задач с использованием большого количества компьютеров, образующих кластер, – ...
- 17) Сокращенное английское название языка структурированных запросов – ...
- 18) Метод исследования, характеризующийся выделением и изучением отдельных частей объектов исследования, – ... (слово происходит от греч. «разложение»)
- 19) Сокращенное английское название семейства нереляционных баз данных – ...
- 20) Установите соответствие баз данных и их характеристик:
- 21) Установите правильный порядок действий при проектировании баз данных:
- 22) Рассматривается несколько технологий систем управления базами данных (СУБД). Это технологии Hadoop, Cassandra, HDInsight и Redis. Какая из данных технологий не использует принцип MapReduce?
- 23) Неверно, что существует такой вид искусственных нейронных сетей, как ...
- 24) Такой вид машинного обучения, как ..., основывается на взаимодействии обучаемой системы со средой
- 25) В изучении нейронных сетей и машинного обучения часто упоминают закон Мура, согласно которому ...
- 26) Градиентом в искусственных нейронных сетях называют ...
- 27) Минимальное количество слоев, которые должна иметь нейронная сеть, для того чтобы к ней возможно было применить алгоритм обратного распространения, – ...
- 28) Процесс привлечения к решению задач в инновационной сфере широкого круга лиц по типу субподрядной работы на добровольных началах с применением информационно-коммуникационных технологий – это ...
- 29) Область исследования, включающая создание компьютерных алгоритмов для преобразования данных в обоснованные действия, – это ... обучение





- 30) Вычислительная единица в искусственных нейронных сетях, которая получает информацию, производит над ней простые вычисления и передает ее дальше, – это ...
- 31) Установите соответствие функций, применяемых для расчета ошибок выходного сигнала нейронной сети, и формул для их вычисления:
- 32) Укажите правильную последовательность шагов алгоритма машинного обучения (ML):
- 33) Вам необходимо принять решение о выдаче кредита клиенту. Какой метод машинного обучения вам в этом поможет?
- 34) Если расстоянием между кластерами называется минимальное из расстояний между парами объектов, один из которых входит в первый кластер, а другой – во второй, то это расстояние ...
- 35) Процедура снижения признакового пространства ...
- 36) Возможность снижения размерности признакового пространства обусловлена ... (укажите 3 варианта ответа)
- 37) Элементы матрицы факторных нагрузок меняются в пределах ...
- 38) Если в результате расчетов получены значения двух первых главных компонент $\lambda_1=1,928$ и $\lambda_2=0,072$, тогда относительный вклад первой главной компоненты в суммарную дисперсию (с точностью до 0,1) равен ... %
- 39) Если был проведен компонентный анализ по шести переменным и суммы квадратов факторных нагрузок для трех главных компонент равны 3,5, 1,3 и 0,3 соответственно, тогда относительный вклад первых двух компонент в общую дисперсию составляет ... %
- 40) Если выборочные данные о потреблении и накоплении пяти домохозяйств представлены матрицей «объект – свойство» (см. ниже), то, считая, что различия в потреблении в 4 раза важнее различий в накоплении, а сумма весовых коэффициентов равна 1, взвешенное евклидово расстояние между домохозяйствами № 1 и № 2 с точностью до 0,01 будет равно ...
- 41) Расстояние Манхэттен между векторами (-6, 3, -2) и (-1, 3, -3) равно ...
- 42) Расположите в правильной последовательности этапы кластерного анализа методом k-средних:
- 43) Установите соответствие между метриками расстояния между объектами и формулами расчета:





- 44) Вашей компании необходимо проанализировать продуктовые наборы для их своевременной закупки и совместного размещения в торговом зале. Какой метод Data Mining вы будете использовать?
- 45) Функция ... реализует алгоритм классификации в языке программирования R
- 46) Функция ... реализует алгоритм кластеризации в языке программирования R
- 47) Библиотека ... является базовой для Data Science на языке программирования Python
- 48) В языке программирования R ошибочными являются такие варианты присвоения, как ... (укажите 2 варианта ответа)
- 49) Язык программирования, который объединяет объектно-ориентированный и функциональный подходы и используется в Apache Spark, – это язык ...
- 50) Способность языка программирования распознавать типы переменных – это ...
- 51) Язык программирования ... имеет лицензию от компании MathWorks и широко используется для вычислений как в научных, так и в промышленных целях
- 52) Язык программирования R является потомком языка ...
- 53) Установите соответствие между аргументами функции leaflet() и их содержанием:
- 54) Укажите правильную последовательность этапов анализа данных в языке программирования R:
- 55) Вам необходимо выполнить интерактивную визуализацию данных. Какой из языков программирования содержит больше библиотек и фреймворков для интерактивной визуализации и лучше подойдет для выполнения этой задачи?
- 56) Velocity в контексте характеристик Big Data означает ...
- 57) Неверно, что Variety в контексте характеристик Big Data означает ... (укажите 3 варианта ответа)
- 58) Данные являются структурированными в ... (укажите 2 варианта ответа)
- 59) Первые большие наборы данных начали использовать ...





- 60 Минимальная единица измерения информации – это ...
- 61 Процесс преобразования данных в код, который можно распознать только с помощью определенного ключа или пароля, – это ...
- 62 ... – это процесс преобразования данных в выводы, на основе которых будут приниматься решения и строиться действия с помощью людей, процессов и технологий
- 63 Целое четырехзначное число, отражающее количество зеттабайт в йоттабайте, – ...
- 64 Установите соответствие между понятием и его содержанием:
- 65 Установите соответствие целей инфраструктуры Hadoop и их характеристик:
- 66 Причина создания баз данных NoSQL – ...
- 67 Неверно, что MapReduce ...
- 68 В реляционной модели представленные данные для пользователя передаются в виде ...
- 69 SQL – это ...
- 70 ... – это группа файлов стандартной структуры, служащих для хранения данных и организованных таким образом, чтобы обеспечить легкий доступ к ним, эффективное управление и обновление
- 71 Метод, используемый для предотвращения дублирования данных в реляционных базах данных, – ...
- 72 Название на английском языке системы или репозитория данных, хранящихся в необработанном формате, – ...
- 73 Сокращенное английское название семейства нереляционных баз данных – ...
- 74 Установите соответствие этапов MapReduce и их содержания:
- 75 Установите правильный порядок действий при проектировании баз данных (БД):
- 76 Установите соответствие баз данных и их характеристик:
- 77 Краудсорсинг – это ...





- 78) Консолидация – это ...
- 79) Интеграция – это ...
- 80) Машинное обучение – это ...
- 81) Фамилия британского ученого, который разработал эмпирический тест, призванный определить способность компьютера мыслить как человек, – ...
- 82) Простейший вид нейронной сети, который представляет собой алгоритм для выполнения двоичной классификации, – это ...
- 83) Вычислительная единица в искусственных нейронных сетях, которая получает информацию, производит над ней простые вычисления и передает ее дальше, – это ...
- 84) Вектор частных производных функции потерь по весам нейронной сети, который указывает на направление наибольшего роста этой функции, – это ...
- 85) Установите соответствие видов искусственных нейронных сетей и их характеристик:
- 86) Определите правильную последовательность этапов распределенных вычислений MapReduce:
- 87) Установите соответствие функций, применяемых для расчета ошибок выходного сигнала нейронной сети, и формул для их вычисления:
- 88) Укажите правильную последовательность шагов алгоритма машинного обучения (ML):
- 89) Web Mining применяет технологию Data Mining для анализа ... (укажите 4 варианта ответа)
- 90) ... является шириной коробки в диаграмме «ящик с усами»
- 91) Data Mining – это процесс обнаружения в сырых данных ... (укажите 3 варианта ответа)
- 92) Деревья классификации относятся к стратегии ...
- 93) Метод Data Mining, который относит объекты к одному из заранее известных классов, – это ...
- 94) Группа элементов, обладающих каким-то общим свойством и находящихся на небольшом расстоянии друг от друга, – это ...





- 95) В числовом ряду 3, 5, 8, 12, 2, 9, 6, 10, 11, 14 медиана (с точностью до 0,1) равна ...
- 96) Если все значения признака увеличить в 16 раз, дисперсия увеличится в ... раз
- 97) Расположите в правильной последовательности этапы дивизимного метода кластерного анализа:
- 98) Установите соответствие между описательными статистиками и их определением:
- 99) Расположите в правильной последовательности этапы кластерного анализа методом k-средних:
- 100) Установите соответствие между метриками расстояния между объектами и формулами расчета:
- 101) Чтобы получить помощь в программной среде R, можно ... (укажите 3 варианта ответа)
- 102) В языке программирования R используются такие типы данных, как ... (укажите 5 вариантов ответа)
- 103) Циклический алгоритм используется при вычислении ... (укажите 2 варианта ответа)
- 104) Разветвляющийся алгоритм – это ...
- 105) Нажатие клавиши «...» предложит возможные варианты продолжения текста кода, включая выбор параметров и переменных в аргументах набираемых команд
- 106) Нажатием сочетания клавиш «...» и «-» можно написать оператор присваивания "
- 107) В языке программирования R тип данных «строка» – это ...
- 108) Язык программирования R является потомком языка ...
- 109) Укажите правильную последовательность этапов анализа данных в языке программирования R:
- 110) Установите соответствие между пакетами языка программирования R и их назначением:
- 111) Установите соответствие между аргументами функции leaflet() и их содержанием:





- 112) В вашей компании необходимо организовать сбор, хранение и обработку больших данных. На что в первую очередь должно обратить внимание руководство при организации хранения и обработки данных?
- 113) Разным группам пользователей с различным уровнем доступа требуется доступ к одной и той же информации. Что следует предпринять руководству?
- 114) Ваша компания использует различные технологии для сбора и хранения данных. Приведите пример данных, когда эффективнее использовать технологию Hadoop.
- 115) Вам необходимо организовать потоковую обработку событий в режиме реального времени. Какую технологию необходимо использовать?
- 116) Вам нужно рассчитать необходимые параметры для создания обшивки самолета. Какая из областей машинного обучения вам в этом поможет?
- 117) Вам необходимо предсказать значения непрерывной числовой величины для входных данных. Какой метод вам в этом поможет?
- 118) Вашей компании необходимо закупить новый вид сырья. Вам необходимо узнать, насколько может упасть стоимость этого сырья в ближайшем будущем и когда это может произойти. Какой метод Data Mining вы будете использовать?
- 119) Вашей компании необходимо определить структуру клиентов по их вкладу в доходы компании. Какой метод Data Mining вы будете использовать?
- 120) Вам необходимо импортировать в программную среду RStudio текстовый файл с разделителем. Какую функцию вы будете использовать?
- 121) Выполнены следующие строки кода в программной среде RStudio:
`vec <- c(3,5,2,1,8,4,9,10,3,15,1,11)`
`vec[2]`

