

По умолчанию для Машинное обучение (09.04.01, 2021)/Проверка остаточных знаний/ПК-26

1. В таблицах с данными признакам соответствуют...

- a. Строки
- b. Столбцы
- c. Одиночные ячейки

В таблицах с данными признакам соответствуют (Множественный выбор / Только один ответ)

2. Значение Мах-метрики для векторов (1,2,3) и (-1,-1,-1) равно

- a. 2
- b. 3
- c. 4

Значение Мах-метрики для векторов (1,2,3) и (-1,-1,-1) (Множественный выбор / Только один ответ)

3. Какой признак является категориальным?

- a. Группа крови у человека
- b. Вес человека
- c. Количество детей у человека

Какой признак является категориальным? (Множественный выбор / Только один ответ)

4. Какой элемент по сути является выбросом в следующей последовательности: 1,0,100,1,99,0,50,1,100,0,99,0?

- a. 50
- b. 100
- c. 0

Какой элемент по сути является выбросом в следующей последовательности (Множественный выбор / Только один ответ)

5. Коэффициент корреляции между вектором (1,2,3) и вектором (1,2,3) равен

- a. 1
- b. 0
- c. -1

Коэффициент корреляции между вектором (1,2,3) и вектором (1,2,3) равен (Множественный выбор / Только один ответ)

6. Коэффициент корреляции никогда не может принять ...

- a. Значение равное 2
- b. Положительное значение
- c. Отрицательное значение

Коэффициент корреляции никогда не может принять (Множественный выбор / Только один ответ)

7. Медиана значений вектора (-1,0,1,1,4,5,6) равна

- a. 1
- b. 4
- c. Медианы существуют только в школьном курсе геометрии

Медиана значений вектора (-1,0,1,1,4,5,6) (Множественный выбор / Только один ответ)

8. Отклонение (среднее квадратическое отклонение) для вектора (1,1,1) равно...

- a. 0
- b. 1
- c. 3

Отклонение (среднее квадратическое отклонение) для вектора (1,1,1) (Множественный выбор / Только один ответ)

9. Среднее значение вектора чисел (0,4,5) равно...

- a. 2
- b. 3
- c. 4

Среднее значение вектора чисел (0,4,5) (Множественный выбор / Только один ответ)

10. Тренировочная выборка обязательно содержит

- a. Точные значения целевого признака
- b. Не менее 10 объектов
- c. Расписание тренировок

Тренировочная выборка обязательно содержит (Множественный выбор / Только один ответ)

По умолчанию для Машинное обучение (09.04.01, 2021)/Проверка остаточных знаний/ПК-27

1. Алгоритм "Наивного Байеса" выдает...

- a. метки класса
- b. вероятности принадлежности к классам
- c. рекомендации по принадлежности к классам

Алгоритм "Наивного Байеса" выдает (Множественный выбор / Только один ответ)

2. Если при решении задачи регрессии используется комитет из трёх моделей, выдавших значения 2,3,7 то окончательным ответом будет значение

- a. 4
- b. 3
- c. 7

Если при решении задачи регрессии используется комитет из трёх моделей (Множественный выбор / Только один ответ)

3. Какая зависимость между целевым признаком Y и нецелевым признаком X никогда не может быть выдана в качестве ответа алгоритмом линейной регрессии?

- a. $Y=X^2$
- b. $Y=100X$
- c. $Y=-X+1$

Какая зависимость между целевым признаком Y и нецелевым признаком X никогда не может быть выдана в качестве ответа алгоритмом линейной регрессии (Множественный выбор / Только один ответ)

4. Какую зависимость восстановит алгоритм регрессии по следующим парам данных? X Y 1 3 2 5 3 7

- a. $Y=2X+1$
- b. $Y=3X$
- c. $Y=X+3$

Какую зависимость восстановит алгоритм регрессии по следующим парам данных (Множественный выбор / Только один ответ)

5. Метод "Изолирующий лес" применяется для...

- a. поиска выбросов
- b. решения задачи предсказания
- c. кластеризации

Метод "Изолирующий лес" (Множественный выбор / Только один ответ)

6. При каком значении константы C алгоритм регуляризации автоматически превратиться в классический алгоритм линейной регрессии?

- a. 0
- b. 1
- c. -1

При каком значении константы C алгоритм регуляризации автоматически превратиться в классический алгоритм линейной регрессии? (Множественный выбор / Только один ответ)

7. При построении дерева решений на каждом этапе используется признак с ... значением неопределенности Джини

- a. минимальным
- b. максимальным
- c. наиболее информативным

При построении дерева решений на каждом этапе используется признак с ... значением неопределенности Джини (Множественный выбор / Только один ответ)

8. Пример алгоритма кластеризации с заранее неизвестным числом кластеров

- a. FOREL
- b. k-means
- c. kNN

Пример алгоритма кластеризации с заранее неизвестным числом кластеров (Множественный выбор / Только один ответ)

9. Пример алгоритма кластеризации с известным заранее числом кластеров

- a. FOREL
- b. k-means
- c. kNN

Пример алгоритма кластеризации с известным заранее числом кластеров (Множественный выбор / Только один ответ)

10. Эффект снижения точности предсказания при увеличении количества признаков называется

- a. Проклятием размерности
- b. Головокружением от успехов
- c. Тепловой смертью

Эффект снижения точности предсказания при увеличении количества признаков называется (Множественный выбор / Только один ответ)

По умолчанию для Машинное обучение (09.04.01, 2021)/Введение в машинное обучение и основные понятия статистики

1. Будем искать выбросы с помощью метрики. Для каждого объекта выборки найдем расстояние до его ближайшего соседа и сформируем таблицу: Какой из объектов больше других похож на выброс?

- a. A
- b. B
- c. C
- d. D
- e. E

Будем искать выбросы с помощью метрики (Множественный выбор)

2. В симметричной выборке...

- a. среднее значение близко к нулю
- b. разница между медианой и средним значением близка к нулю
- c. отклонение близко к нулю

В симметричной выборке (Множественный выбор / Только один ответ)

3. Однажды я попросил, чтобы студенты ответили на 2 вопроса анкеты «ваш год рождения» и «ваш возраст». Из их ответов я сформировал таблицу, в которой был столбец P = «год рождения студента» и Q = «возраст студента». Коэффициент корреляции признаков P , Q

- a. положительный
- b. отрицательный

Однажды я попросил, чтобы студенты ответили на 2 вопроса анкеты (Множественный выбор / Только один ответ)

4. Первая и третья квартиль значений признака P равны 2, 4 соответственно. Какие из следующих значений будут считаться выбросами?

- a. -1,5

- b. 0,5
- c. 0
- d. 3
- e. 6,5
- f. 7,5
- g. 8

Первая и третья квартиль значений признака Р равны 2, 4 соответственно (Множественный выбор)

5. Представьте, что у вас есть выборка значений некоторого признака Р, и среднее значение в ней равно 50. Для этой выборки был критерий Шовене поиска выбросов. Оказалось, что элементы выборки равные 42 и 57 были признаны выбросами. Какие из указанных ниже элементов выборки критерий Шовене также определит как выбросы?

- a. 43
- b. 58
- c. 56
- d. 4

Представьте, что у вас есть выборка значений некоторого признака Р, (Множественный выбор)

6. В таблице приведены сведения о выставленных на продажу квартирах. Признак «материал стен» является ...

- a. числовым
- b. номинальным
- c. порядковым
- d. бинарным

Признак «материал стен» является (Множественный выбор / Только один ответ)

7. В таблице приведены сведения о выставленных на продажу квартирах. Признак «наличие лифта» является ...

- a. числовым
- b. номинальным
- c. порядковым
- d. бинарным

Признак «наличие лифта» является (Множественный выбор / Только один ответ)

8. В таблице приведены сведения о выставленных на продажу квартирах. Признак «площадь» является ...

- a. числовым
- b. номинальным
- c. порядковым
- d. бинарным

Признак «площадь» является (Множественный выбор / Только один ответ)

9. Среднее значение, отклонение и медиана 10 значений признака Р равны 10, 1,1, 9 соответственно. Проверьте выборку на симметричность. Какие из следующих значений будут выбросами?

- a. 6
- b. 6,5

- c. 7
- d. 13
- e. 13,5
- f. 14

Среднее значение, отклонение и медиана 10 значений признака P равны 10 1,1 9 соответственно (Множественный выбор)

10. Однажды я попросил, чтобы студенты ответили на 2 вопроса анкеты «ваш год рождения» и «ваш возраст». Из их ответов я сформировал таблицу, в которой был столбец P=«год рождения студента» и Q=«возраст студента». Оказывается, значение коэффициента корреляции признаков P,Q зависит от месяца, в котором проводилось анкетирование (это не шутка!). Укажите два месяца, которым соответствует наименьшее (по модулю) значение коэффициента корреляции признаков P,Q.

- a. январь
- b. февраль
- c. март
- d. апрель
- e. май
- f. июнь
- g. июль
- h. август
- i. сентябрь
- j. октябрь
- k. ноябрь
- l. декабрь

Укажите два месяца, которым соответствует наименьшее (по модулю) значение коэффициента корреляции (Множественный выбор)

11. Дана таблица с пропущенным значением признака P3 для последнего объекта. Восстановим пропущенное значение. Если пропущенное значение заменить на медиану признака P3, то оно будет равно

16. Дана таблица с пропущенным значением признака P3 для последнего объекта. Если пропущенное значение заменить на медиану признака (Числовой ответ)

12. Дан вектор значений признака $P=(1,0,5,2,2)$. Нормализуем этот вектор по формуле, использующей минимальное и максимальное значение признака P. Значение первой координаты нормализованного вектора будет равно

Дан вектор значений признака $P=(1,0,5,2,2)$. Значение первой координаты (Числовой ответ)

13. Дан вектор значений признака $P=(1,0,5,2,2)$. Нормализуем этот вектор по формуле, использующей среднее значение и отклонение признака P. Значение последней координаты нормализованного вектора будет равно

Ответ: 0

Дан вектор значений признака $P=(1,0,5,2,2)$. Значение последней координаты (Числовой ответ)

- 14. Дана таблица с объектами и их признаками P_1, P_2 . Чему равен коэффициент корреляции признаков P_1 и P_2 ? Ответ округлить до двух знаков после запятой.**

Дана таблица с объектами и их признаками P_1, P_2 (Числовой ответ)

- 15. Дана таблица с пропущенным значением признака P_3 для последнего объекта. Восстановим пропущенное значение. Если пропущенное значение заменить на среднее значение признака P_3 , то оно будет равно**

Дана таблица с пропущенным значением признака P_3 для последнего объекта (Числовой ответ)

- 16. Дана таблица. Значение последнего признака у последнего объекта пропущено. Восстановим его с помощью tanh -метрики (без проведения нормализации). Восстановленное значение равно**

Дана таблица. Значение последнего признака у последнего объекта пропущено (Числовой ответ)

- 17. Даны два вектора значений признаков $P_1=(0,1,2)$ $P_2=(2,1,0)$. Евклидово расстояние между ними равно (ответ округлить до двух знаков после запятой)**

Даны два вектора значений признаков $P_1=(0,1,2)$ $P_2=(2,1,0)$. Евклидово расстояние (Числовой ответ)

- 18. Даны два вектора значений признаков $P_1=(0,1,2)$ $P_2=(2,1,0)$. Расстояние в метрике Манхэттен между ними равно**

Даны два вектора значений признаков $P_1=(0,1,2)$ $P_2=(2,1,0)$. Расстояние в метрике Манхэттен (Числовой ответ)

- 19. Даны два вектора значений признаков $P_1=(0,1,2)$ $P_2=(2,1,0)$. Расстояние между ними в tanh -метрике равно**

Даны два вектора значений признаков $P_1=(0,1,2)$ $P_2=(2,1,0)$. Расстояние между ними в tanh -метрике (Числовой ответ)

20. На основании опроса 40 мужчин и 60 женщин был составлен вектор P , состоящий из 100 нулей и единиц, причем i -ая координата P равна 1 (0), если i -й респондент был мужчиной (женщиной). Чему равна медиана признака P ?

На основании опроса 40 мужчин и 60 женщин был составлен вектор P . Чему равна медиана признака P ? (Числовой ответ)

21. На основании опроса 40 мужчин и 60 женщин был составлен вектор P , состоящий из 100 нулей и единиц, причем i -ая координата P равна 1 (0), если i -й респондент был мужчиной (женщиной). Чему равна мода признака P ?

На основании опроса 40 мужчин и 60 женщин был составлен вектор P . Чему равна мода признака P ? (Числовой ответ)

22. На основании опроса 40 мужчин и 60 женщин был составлен вектор P , состоящий из 100 нулей и единиц, причем i -ая координата P равна 1 (0), если i -й респондент был мужчиной (женщиной). Чему равно среднее значение P ?

На основании опроса 40 мужчин и 60 женщин был составлен вектор P . Чему равно среднее значение P ? (Числовой ответ)

23. Отклонение значений признака $P=(1,1,1,1,1,1)$ равно

Отклонение значений признака $P=(1,1,1,1,1,1)$ (Числовой ответ)

24. Первая квартиль выборки - это такое число, что ровно _____ процентов значений этой выборки больше (!) него.

Первая квартиль выборки (Числовой ответ)

25. Следующая таблица содержит информацию об оценках, выставленных фильмам Требуется оценить, какую оценку поставит Саша фильму «Гарри Поттер». Сделаем это с помощью метрики Манхэттен (для простоты вычислений данные в таблице нормировать не нужно). Для этого подсчитаем расстояния от Саши до других людей, используя информацию из первых 3 столбцов, а далее получим ожидаемую оценку для «Гарри Поттера» равную _____ (округлить до одного знака после запятой).

Следующая таблица содержит информацию об оценках, выставленных фильмам (Числовой ответ)

26. Таблица содержит сведения о заказах анонимных пользователей в интернет-магазине. По строкам отложены наименования товаров, а столбцы — номера заказов (в ячейке стоит 1, если товар входит в соответствующий заказ; 0 — в противном случае). При построении рекомендательной системы необходимо оценить степень схожести товаров с помощью вычисления евклидовой метрики. Наиболее похожим (близким) на товар А будет товар С, и расстояние между этими товарами равно _____ (ответ округлить до сотых).

Таблица содержит сведения о заказах анонимных пользователей в интернет-магазине (Числовой ответ)

27. Третья квартиль выборки - это такое число, что ровно _____ процентов значений этой выборки меньше него.

Третья квартиль выборки (Числовой ответ)

28. Даны значения признака Р для шести объектов выборки $P = (-1, 0, 4, 2, 1, 2)$. Чему равна медиана признака Р?

Чему равна медиана признака Р? (Числовой ответ)

29. Даны значения признака Р для шести объектов выборки $P = (-1, 0, 4, 2, 1, 2)$. Чему равно отклонение признака Р? Ответ округлить до двух знаков после запятой.

Чему равно отклонение признака Р? (Числовой ответ)

30. Даны значения признака Р для шести объектов выборки $P = (-1, 0, 4, 2, 1, 2)$. Чему равно среднее значение признака Р? Ответ округлить до двух знаков после запятой.

Чему равно среднее значение признака Р? (Числовой ответ)

31. Дана таблица с объектами. В данной задаче выбросы будем искать по следующему правилу: «Выбросом будет считаться объект, у которого суммарное расстояние от него до остальных объектов выборки наибольшее». Какой объект в указанной таблице будет выбросом? (при вычислении использовать метрику Манхэттен, нормализацию не проводить)

Дана таблица с объектами. В данной задаче выбросы будем искать по следующему правилу (Короткий ответ)

По умолчанию для Машинное обучение (09.04.01, 2021)/Кластеризация

1. Дана таблица с объектами. Какие объекты попадают в окружность с центром (1,2) и радиусом 2,5 (используется max-метрика)?
- a. A
 - b. B
 - c. C
 - d. D
 - e. E

Дана таблица с объектами (Множественный выбор)

2. Дана таблица с объектами. Планируется разбить объекты на два кластера. На первой итерации работы алгоритма k-means были выбраны точки (2,3) и (1,1). Какие объекты будут отнесены после первой итерации алгоритма к кластеру, определяемому первой точкой (используется метрика Манхэттен)?
- a. A
 - b. B
 - c. C
 - d. D
 - e. E

Дана таблица с объектами (Множественный выбор)

3. Набор объектов A, B, C, D, E, F, G планируется разбить на 3 кластера. Для этого был применен графовый алгоритм, и было построено остовное дерево минимального веса, состоящее из ребер AE, BD, CD, DF, ED, EG. Длины ребер соответственно равны 1, 2, 4, 1, 3, 2. Какие объекты попадут в один кластер с объектом A?
- a. B
 - b. C
 - c. D
 - d. E
 - e. F
 - f. G

Набор объектов A,B,C,D,E,F,G планируется разбить на 3 кластера (Множественный выбор)

4. Расстояния между объектами в некоторой метрике приведены в таблице. Представим данные в виде графа и запустим алгоритм кластеризации, в котором число кластеров заранее не задано. Входной параметр алгоритма равен $R=2,5$. Какие из объектов попадут в один кластер с объектом A?
- a. B
 - b. C
 - c. D
 - d. E

Расстояния между объектами в некоторой метрике (Множественный выбор / Только один ответ)

5. Если кластер состоит лишь из одного объекта, то расстояние от этого объекта до центра кластера равно...

Если кластер состоит лишь из одного объекта (Числовой ответ)

6. Пусть объекты из следующей таблицы образуют один кластер. Укажите координаты центра этого кластера (нужно указать три числа через пробел).

Пусть объекты из следующей таблицы образуют один кластер (Короткий ответ)

По умолчанию для Машинное обучение (09.04.01, 2021)/Линейная регрессия

1. Дана таблица, содержащая истинное значение целевого признака Y и предсказанное значение Y' . Значение MAE равно (ответ в виде десятичной дроби). {#1} Значение MAPE равно (количество процентов округлить до целого). {#2}

- {1:SHORTANSWER:=0,9}
- {1:SHORTANSWER:=33}

Дана таблица, содержащая истинное значение целевого признака Y (Вложенные ответы (Cloze))

2. Объекты в следующей таблице имеют два нецелевых признака X_1, X_2 . Проблема в том, что существует сильная (даже линейная) зависимость между признаками X_1, X_2 . Построим модель линейной регрессии с регуляризацией для предсказания признака Y (значение константы регуляризации C положить равным 1). Полученная модель будет иметь вид: $Y = \{#1\} X_1 + \{#2\} X_2 + \{#3\}$ (найденные значения округлить до двух знаков после запятой).

- {1:SHORTANSWER:=0,62}
- {1:SHORTANSWER:=-0,1}
- {1:SHORTANSWER:=0,17}

Объекты в следующей таблице имеют два нецелевых признака X_1, X_2 (Вложенные ответы (Cloze))

3. Объекты в следующей таблице имеют один нецелевой признак X . Построим модель линейно регрессии для предсказания признака Y . Полученная модель будет иметь вид: $Y = \{#1\} X + \{#2\}$

- {1:SHORTANSWER:=0,8}
- {1:SHORTANSWER:=-0,2}

Объекты в следующей таблице имеют один нецелевой признак X (Вложенные ответы (Cloze))

4. Какие существуют проблемы, связанные с построением модели линейной регрессии?

- a. мультиколлинеарность
- b. бесконечное множество решений у системы уравнений, которая возникает при нахождении весов модели
- c. большие по модулю значения весов

Какие существуют проблемы, связанные с построением модели линейной регрессии? (Множественный выбор)

5. Укажите известные вам модификации модели линейной регрессии

- lasso
- a. регуляризация
- b. кластеризация
- c. релинеаризация

Укажите известные вам модификации модели линейной регрессии (Множественный выбор)

6. Чему равна частная производная функции $f(x,y)=2x^2-3xy+4y^2-2x+y+10$ по переменной x ?

- $2x-3y-2$
- a. $4x-3y-2$
- b. $4x-3+y$
- c. $x-2$
- d. $4x-3xy+8y+9$

Чему равна частная производная функции (Множественный выбор / Только один ответ)

По умолчанию для Машинное обучение (09.04.01, 2021)/Задача классификации

1. Дана таблица с объектами. Построим линейный классификатор, в котором выражение $[Mi < 0]$ мажорируется функцией $(1-M)^2$. Получим следующее правило классификации: если $X > \{ \#1 \}$, то объект из класса 1, иначе объект из класса -1 (ответ записать в виде десятичной дроби, округленной до двух знаков после запятой).

- {1:SHORTANSWER:=-0,33}

Вопрос 8 (Вложенные ответы (Cloze))

2. Дана матрица ошибок, построенная по результатам работы некоторого алгоритма классификации. Общая точность (accuracy) равна $\{ \#1 \}$, precision = $\{ \#2 \}$, recall = $\{ \#3 \}$

- {1:SHORTANSWER:=0,57}
- {1:SHORTANSWER:=0,6}
- {1:SHORTANSWER:=0,43}

Дана матрица ошибок, построенная по результатам работы некоторого алгоритма классификации (Вложенные ответы (Cloze))

3. Дана функция $f(u,v)=u^2-2v^2-2u+4v+1$. Ее градиент в точке $(0,0)$ равен $(\{ \#1 \}, \{ \#2 \})$. Будем искать точку минимума с помощью метода градиентного спуска. Длину шага полагаем равной $0,5$. В какую точку мы попадем из точки $(0,0)$ после одной итерации метода градиентного спуска? Ответ: $(\{ \#3 \}, \{ \#4 \})$.

- $\{1:SHORTANSWER:=-2\}$
- $\{1:SHORTANSWER:=4\}$
- $\{1:SHORTANSWER:=1\}$
- $\{1:SHORTANSWER:=-2\}$

Дана функция $f(u,v)$ (Вложенные ответы (Cloze))

4. Даны объекты тренировочной выборки. На вход алгоритму kNN подается объект F с признаками $X_1=0, X_2=0$. При $k=3$ объект F будет отнесен к классу $\{ \#1 \}$. При $k=5$ объект F будет отнесен к классу $\{ \#2 \}$. При решении задачи использовать евклидову метрику, нормализацию данных не проводить.

- $\{1:SHORTANSWER:=0\}$
- $\{1:SHORTANSWER:=1\}$

Даны объекты тренировочной выборки. На вход алгоритму kNN подается объект F (Вложенные ответы (Cloze))

5. Один очень тупой классификатор C относит все объекты к классу 1. Допустим, что выборка состоит из 50 объектов: 20 из них действительно принадлежат классу 0, а 30 из них действительно принадлежат классу 1. Каждый из объектов выборки был подан классификатору C, в результате работы которого была заполнена матрица ошибок, в которой общая точность (accuracy) равна $\{ \#1 \}$, $precision = \{ \#2 \}$, $recall = \{ \#3 \}$

- $\{1:SHORTANSWER:=0,6\}$
- $\{1:SHORTANSWER:=0,6\}$
- $\{1:SHORTANSWER:=1\}$

Один очень тупой классификатор C относит все объекты к классу 1 (Вложенные ответы (Cloze))

6. Заполните пропуск. По тренировочной выборке из 70 элементов был построен некоторый классификатор. Мы взяли и проверили его качество на тестовой выборке, состоящей из 30 элементов. Сумма чисел $TP+FP+FN+TN$ из матрицы ошибок равна $\{ \#1 \}$

- $\{1:SHORTANSWER:=30\}$

По тренировочной выборке из 70 элементов был построен некоторый классификатор (Вложенные ответы (Cloze))

7. Выбор оптимального признака для дерева можно осуществить с помощью таких понятий как...

- a. Неопределенность Джини
- b. Энтропия
- c. Статистическая информативность

Выбор оптимального признака для дерева можно осуществить (Множественный выбор)

8. Дана таблица с объектами. Вспомните суть метода SVM и, не проводя вычислений, попытайтесь угадать правило, если $a \cdot X_1 + b \cdot X_2 + c > 0$, то объект принадлежит классу 1 (иначе объект принадлежит классу -1)

- a=1 b=0 c=1
- a. a=0 b=1 c=1,5
- b. a=1, b=1, c=-1,5
- c. a=1 b=1 c=1
- d. a=1,5 b=-1 c=1,5
- e. a=1,5 b=1 c=-1

Дана таблица с объектами (Множественный выбор / Только один ответ)

9. Дана таблица с объектами. Каждый объект имеет 2 нецелевых признака и известную метку класса. Линейный классификатор имеет вид: если $a \cdot X_1 + b \cdot X_2 + c > 0$, то объект принадлежит классу 1 (иначе объект принадлежит классу -1). Какими должны быть значения весов a, b, c, чтобы классификатор правильно классифицировал все объекты в таблице?

- a=1, b=1, c=-1
- a. a=1, b=1, c=1
- b. a=1, b=-1, c=0
- c. a=0, b=1, c=1
- d. a=1, b=0, c=0
- e. a=0, b=1, c=0

Дана таблица с объектами. Каждый объект имеет 2 нецелевых признака (Множественный выбор / Только один ответ)

10. Искусственный нейрон имеет 2 входа с весами 1, 2 соответственно. Ко входному сигналу прибавляется число 1 и применяется сигмоидная функция. Чему будет равно значение на выходе из нейрона, если на первый вход было подано число 1, а на второй вход было подано число -1?

- 1
- a. -0,5
- b. 0
- c. 0,5
- d. 1
- e. 1,5
- f. 2

Искусственный нейрон имеет 2 входа (Множественный выбор / Только один ответ)

11. Модель «изолирующий лес» предназначена для:

- предсказания метки класса
- a. поиска выбросов
- b. кластеризации
- c. предсказания значения числового признака

Модель «изолирующий лес» предназначена (Множественный выбор / Только один ответ)

12. Укажите все величины, отвечающие за качество алгоритмов классификации:

полнота

- a. F-значение
- b. MAE
- c. MAPE
- d. точность

Укажите все величины, отвечающие за качество алгоритмов классификации (Множественный выбор)

13. Чем больше значения FP и FN в матрице ошибок, тем

хуже

- a. лучше

Чем больше значения FP и FN в матрице ошибок (Множественный выбор / Только один ответ)

14. Чем меньше значения TP и TN в матрице ошибок, тем

лучше

- a. хуже

Чем меньше значения TP и TN в матрице ошибок (Множественный выбор / Только один ответ)

15. Чему равен градиент функции $f(u,v)=u*v+1$ в точке (1,2)?

3

- a. (1,2)
- b. (2,1)
- c. 1
- d. (1,1)
- e. (2,2)

Чему равен градиент функции $f(u,v)$ (Множественный выбор / Только один ответ)

16. Дана таблица, объекты в которой обладают нецелевыми признаками P1,P2,P3 и целевым признаком Y Как известно в качестве условия ветвления выбирается признак с минимальным значением неопределенности Джини. Вычислите эту величину с точностью до двух знаков после запятой для признака P1.

Дана таблица, объекты в которой обладают нецелевыми признаками P1,P2,P3 (Числовой ответ)

17. Дана таблица, объекты в которой обладают нецелевыми признаками P1,P2,P3 и целевым признаком Y Как известно в качестве условия ветвления выбирается признак с минимальным значением неопределенности Джини. Вычислите эту величину с точностью до двух знаков после запятой для признака P2.

Дана таблица, объекты в которой обладают нецелевыми признаками P1,P2,P3 (копия) (Числовой ответ)

18. Дана таблица, объекты в которой обладают нецелевыми признаками P_1, P_2, P_3 и целевым признаком Y . Как известно в качестве условия ветвления выбирается признак с минимальным значением неопределенности Джини. Вычислите эту величину для признака P_3 .

- 0
- 0

Дана таблица, объекты в которой обладают нецелевыми признаками P_1, P_2, P_3 (копия) (копия) (Числовой ответ)

19. Модель «случайный лес» состоит из трёх деревьев: Объект N будет отнесен к классу

Модель «случайный лес» состоит из трёх деревьев (Числовой ответ)

20. По тренировочной выборке было построено дерево (в листьях дана информация о количестве объектов тренировочной выборки, попавших в этот лист). Объект с пропусками данных будет отнесен к классу ...

По тренировочной выборке было построено дерево (Числовой ответ)

21. Представим, что было построено дерево решений, и в некоторый лист попало 5 объектов тренировочной выборки. Причем 2 объекта из них принадлежат классу 1. Какую метку дерево будет предсказывать для каждого тестируемого объекта, попавшего в этот лист?

Представим, что было построено дерево решений (Числовой ответ)

По умолчанию для Машинное обучение (09.04.01, 2021)/Вероятностные алгоритмы. Наивный Байес

1. Будем кидать игральный кубик. Вероятность события $A = \{\text{выпало нечетное число}\}$ равна $\{ \#1 \}$. $B = \{\text{выпало число строго больше 4}\}$ равна $\{ \#2 \}$ (ответ округлить до двух знаков после запятой). $C = \{\text{выпало число не меньше 2}\}$ равна $\{ \#3 \}$ (ответ округлить до двух знаков после запятой). Условные вероятности равны $Pr(A|B) = \{ \#4 \}$ $Pr(B|A) = \{ \#5 \}$ $Pr(A|C) = \{ \#6 \}$ $Pr(C|A) = \{ \#7 \}$

- $\{1:SHORTANSWER:=0,5\}$
- $\{1:SHORTANSWER:=0,33\}$
- $\{1:SHORTANSWER:=0,83\}$

- {1:SHORTANSWER:=0,5}
- {1:SHORTANSWER:=0,33}
- {1:SHORTANSWER:=0,4}
- {1:SHORTANSWER:=0,67}

Будем кидать игральный кубик (Вложенные ответы (Cloze))

2. Выборка состоит из 100 объектов, 70 из них принадлежат классу 1, остальные принадлежат классу 0. Известно, что нецелевой бинарный признак P равен 1 для 10% объектов класса 1 и для 50% объектов из класса 0. Рассмотрим новый объект, у которого признак P равен 1. Вероятность, что этот объект принадлежит классу 1, равна {#1}

- {1:SHORTANSWER:=0,32}

Выборка состоит из 100 объектов (Вложенные ответы (Cloze))

3. Дана таблица с объектами. По этой выборке построим наивный байесовский классификатор, который для нового объекта предскажет вероятность принадлежности классу 1 равную {#1}

- {1:SHORTANSWER:=0}

Дана таблица с объектами (Вложенные ответы (Cloze))

4. Для объектов из таблицы с помощью некоторого алгоритма были получены вероятности принадлежности классу 1. Истинные метки классов объектов также известны. Площадь под ROC-кривой (величина AUC) равна {#1}

- {1:SHORTANSWER:=1}

Для объектов из таблицы с помощью некоторого алгоритма (Вложенные ответы (Cloze))

5. Где здесь формула Байеса?

- a. $\Pr(B|A)=\Pr(A|B)*\Pr(B)/\Pr(A)$
- b. $\Pr(B|A)=\Pr(A|B)*\Pr(A)/\Pr(B)$
- c. $\Pr(B|A)=\Pr(A|B)*\Pr(B)/\Pr(A)$
- d. $\Pr(B|A)=\Pr(A|B)/\Pr(B)*\Pr(A)$

Где здесь формула Байеса? (Множественный выбор / Только один ответ)

6. События A, B называются независимыми, если:

- $\Pr(A|B)=\Pr(A)*\Pr(B)$
- a. $\Pr(A|B)=\Pr(B|A)$
- b. $\Pr(AB)=\Pr(A)+\Pr(B)$
- c. $\Pr(A)*\Pr(B)=\Pr(AB)$

События A, B называются независимыми (Множественный выбор / Только один ответ)

-
7. Для объектов из таблицы с помощью некоторого алгоритма были получены вероятности принадлежности классу 1. Истинные метки классов объектов также известны. Площадь под ROC-кривой (величина AUC) равна _____ (ответ округлить до двух знаков после запятой).

Для объектов из таблицы с помощью некоторого алгоритма были получены вероятности принадлежности (Числовой ответ)

По умолчанию для Машинное обучение (09.04.01, 2021)/Отбор оптимального числа признаков и ансамбли алгоритмов

1. Пусть объекты имеют нецелевые признаки P1, P2, P3 и целевой признак Y. Будем использовать генетический алгоритм для отбора признаков для задачи предсказания признака Y. Каким вектором задается «особь», утверждающая, что признаки P1 и P3 информативны, а признак P2 — нет. ($\{ \#1 \}$, $\{ \#2 \}$, $\{ \#3 \}$)

- $\{1:SHORTANSWER:=1\}$
- $\{1:SHORTANSWER:=0\}$
- $\{1:SHORTANSWER:=1\}$

Пусть объекты имеют нецелевые признаки P1, P2, P3 и целевой признак Y (Вложенные ответы (Cloze))

-
2. Рассмотрим работу SMOTE-алгоритма по синтезу новых объектов. Даны два объекта. Параметр линейной комбинации α равен 0,25. Значения признаков синтетического объекта равны: $P1 = \{ \#1 \}$ $P2 = \{ \#2 \}$

- $\{1:SHORTANSWER:=4\}$
- $\{1:SHORTANSWER:=8\}$

Рассмотрим работу SMOTE-алгоритма по синтезу новых объектов (Вложенные ответы (Cloze))

-
3. Была построена модель регрессии, предсказывающая целевой признак Y, и предсказанные значения признаков Y' указаны в таблице. Будем улучшать качество регрессии с помощью процедуры градиентного бустинга. Тогда следующая модель регрессии будет строиться с помощью нового столбца, состоящего из чисел:

- a. $\begin{matrix} 2 & 2 \\ b. & 1 & 2 \\ c. & -1 & 2 \\ d. & 1 & 1 \\ e. & -1 & -1 \\ f. & -2 & -2 \end{matrix}$

Была построена модель регрессии, предсказывающая целевой признак Y (Множественный выбор / Только один ответ)

4. В таблице дана точность модели построенной с помощью нецелевых признаков U, V, W, X (если в таблице стоит 0, то соответствующий признак не использовался при построении модели). Используя жадный алгоритм отбора признаков (который начинает свою работу с пустого набора признаков), найдем все признаки, попавшие в оптимальный набор.

U

- a. W
- b. V
- c. X

В таблице дана точность модели построенной с помощью нецелевых признаков (Множественный выбор)

5. В таблице дана точность модели построенной с помощью нецелевых признаков U, V, W, X (если в таблице стоит 0, то соответствующий признак не использовался при построении модели). Используя жадный алгоритм отбора признаков (который начинает свою работу с пустого набора признаков), найдем все признаки, попавшие в оптимальный набор.

U

- a. V
- b. W
- c. X

В таблице дана точность модели построенной с помощью нецелевых признаков (Множественный выбор)

6. Вспомните работу алгоритма Random Forest (случайный лес). К какому типу ансамбля он относится?

голосование по большинству (комитет)

- a. взвешенное голосование
- b. градиентный бустинг

Вспомните работу алгоритма Random Forest (случайный лес) (Множественный выбор / Только один ответ)

7. Если к модели линейной регрессии применить процедуру бустинга, причем на втором этапе будет также строится линейная модель, то точность предсказания возрастет

- a. уменьшится
- b. останется прежней

Если к модели линейной регрессии применить процедуру бустинга (Множественный выбор / Только один ответ)

8. Пусть объекты имеют нецелевые признаки P_1, P_2, P_3, P_4 и целевой признак Y . Будем использовать генетический алгоритм для отбора признаков для задачи предсказания признака Y . Какое «потомство» произведут «особи» $(1,0,1,0)$, $(0,0,1,1)$ при наборе индексов $\{3,4\}$

$(1,0,0,1)$

- a. $(1,1,1,1)$ и $(0,1,1,0)$
- b. $(1,0,1,1)$ и $(0,0,1,0)$
- c. $(1,1,1,1)$
- d. $(1,0,1,0)$ и $(0,0,1,1)$

Пусть объекты имеют нецелевые признаки P_1, P_2, P_3, P_4 и целевой признак Y (Множественный выбор / Только один ответ)

9. Укажите необходимые условия возникновения проклятия размерности:

используются данные без нормировки

- a. признаков очень много
- b. используются метрические алгоритмы
- c. признаки малоинформативны
- d. существует зависимость между признаками

Укажите необходимые условия возникновения проклятия размерности (Множественный выбор)

10. Для классификации используется схема взвешенного голосования. Допустим, что объект A был определен первым и третьим классификаторами как объект из класса 1, а второй классификатор определил A как объект класса -1. Результат взвешенного голосования, если веса при классификаторах равны 2, 4, 3 соответственно, отнесет объект A к классу

Для классификации используется схема взвешенного голосования (Числовой ответ)

11. Комитет состоит из трех классификаторов. Окончательное решение принимается по большинству голосов. Классификаторы имеют точность 0,8 0,9 1 соответственно. Точность ансамбля алгоритмов равна

Комитет состоит из трех классификаторов (Числовой ответ)

12. Комитет состоит из трех классификаторов. Окончательное решение принимается по большинству голосов. Классификаторы имеют точность 0,8 0,9 0,9 соответственно. Кроме того, второй и третий классификаторы полностью идентичны, то есть их ответы для любого классифицируемого объекта совпадают.

Комитет состоит из трех классификаторов (Числовой ответ)
