

## Вопросы для аналитиков и специалистов по машинному обучению.

1. Какой из перечисленных ниже функционалов качества используются в задачах классификации и является дифференцируемым?

- ☐  $Q(w) = \sum_{i=1}^l |a(x_i, w) - y_i|$
- ☐  $Q(w) = \sum_{i=1}^l | \text{sign}(a(x_i, w)) - y_i |$
- ☐  $Q(w) = \sum_{i=1}^l (a(x_i, w) - y_i)^2$
- ☒  $Q(w) = \sum_{i=1}^l y_i \ln(a(x_i, w)) + (1 - y_i) \ln(1 - a(x_i, w))$

2. Почему в градиентном спуске на каждой итерации делается шаг в сторону антиградиента?

- ☐ Антиградиент функционала ошибки зависит только от одного объекта.
- ☐ Антиградиент легко найти, в отличие от других направлений.
- ☒ Антиградиент совпадает с направлением наискорейшего убывания.

Правильно

Таким образом ошибка на выборке на каждой итерации минимизируется.

3. Градиент какой функции/функционала и по какому аргументу используется в градиентном спуске при обучении линейной регрессии?

- ☐ Алгоритма — то есть скалярного произведения вектора признаков — по вектору весов.
- ☐ Функционала ошибки — например, среднеквадратичной ошибки — по вектору весов.

Правильно

Такой градиент показывает в какую сторону нужно сдвинуть веса, чтобы уменьшить ошибку на выборке.

- ☐ Функционала ошибки — например, среднеквадратичной ошибки — по прогнозам алгоритма.

4. В чём заключается отличие градиентного спуска от стохастического градиентного спуска?

- ☐ В стохастическом градиентном спуске на каждой итерации делается шаг в случайном направлении.
- ☐ В стохастическом градиентном спуске на каждой итерации используется лишь одно слагаемое в функционале ошибки.
- ☐ В стохастическом градиентном спуске на каждой итерации к антиградиенту добавляется нормальный шум.

5. Для чего при обучении линейных классификаторов используются верхние оценки на пороговую функцию потерь?

- ☐ Чтобы перейти к функции потерь, на которой градиентный спуск будет быстрее сходиться.
- ☐ Чтобы перейти к функции потерь, которая будет легко вычислима — без этого градиентная оптимизация будет занимать слишком много времени.
- ☐ Чтобы заменить разрывную пороговую функцию потерь на гладкую функцию — без этого градиентная оптимизация будет невозможна.

6. Предположим, вы решаете задачу восстановления регрессии с использованием некоторого функционала ошибки  $Q(w, X)$ . Как будет выглядеть новый функционал при добавлении  $L_2$ -регуляризатора с коэффициентом регуляризации  $\lambda$ ?

- ☒  $Q(w, X) + \lambda \sum_{j=1}^d w_j^2$

Правильно

Добавочное слагаемое является квадратом  $L_2$  нормы весов линейной модели.

- ☐  $Q(w, X) + \lambda \sum_{j=1}^d w_j$
- ☐  $Q(w, X) + \lambda \sum_{j=1}^d |w_j|$
- ☐  $Q(w, X) + \lambda \sum_{j=1}^d \frac{1}{w_j^2}$

7. Метод опорных векторов (SVM) — это линейный классификатор, использующий:

☐ Кусочно-линейную функцию потерь (hinge loss) и  $L_2$  регуляризатор.

Правильно

☐ Логистическую функцию потерь и  $L_1$  регуляризатор.

☐ Экспоненциальную функцию потерь и  $L_2$  регуляризатор.

☐ Квадратичную функцию потерь и  $L_1$  регуляризатор.

☐ Логистическую функцию потерь и  $L_2$  регуляризатор.

8. Выберите предположения, выполнение которых необходимо и достаточно для того, чтобы метод наименьших квадратов давал несмещённые и состоятельные оценки истинных коэффициентов регрессии.

☒ Полнота ранга  $X$ : ни один из признаков не является линейной комбинацией других признаков.

Правильно

☒ Линейность отклика:  $y = X\beta + \varepsilon$

Правильно

☐ Гомоскедастичность ошибок

правильно, этот вариант не должен быть выбран

☒

Нормальность ошибок:  $\varepsilon | x \sim N(0, \sigma^2)$

правильно, этот вариант не должен быть выбран

☒

Случайность ошибок:  $E(\varepsilon | x) = 0$

Правильно

☒

Случайность выборки: наблюдения  $(x_i, y_i)$  независимы

Правильно

9. Предположим, вы оцениваете качество работы алгоритма при помощи кросс-валидации с разбиением на  $k$  блоков. Сколько раз будет проведено обучение модели?

- ☐  $\frac{k(k-1)}{2}$
- ☒  $k$
- ☐  $k^2$
- ☐ 1

10. К какому из указанных чисел будет ближе значение метрики AUC-ROC для алгоритма, возвращающего случайный ответ для любого объекта?

- ☐ Невозможно указать точное значение без информации о выборке.
- ☐ 0
- ☐ 0.33
- ☒ 0.5
- ☐ 1

11. Зачем нужно масштабировать признаки перед обучением линейной модели?

☐ Линейная модель не будет иметь смысла при обучении на признаках с разным масштабом.

правильно, этот вариант не должен быть выбран

☒ Градиентный спуск может очень медленно сходиться при разном масштабе признаков.

Правильно

☐ Благодаря масштабированию выборка будет занимать меньше места в оперативной памяти, что позволит ускорить процесс обучения.

12. Выберите верные утверждения про бинарное кодирование категориальных признаков.

☐ Бинарное кодирование категориальных признаков является способом регуляризации линейных моделей.

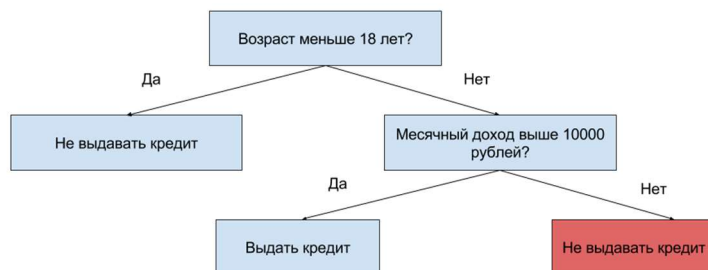
правильно, этот вариант не должен быть выбран

☒ Если обучить над бинарным кодированием линейную модель, то получится, что каждому значению исходного категориального признака будет соответствовать свой вес.

Правильно

☒ Для кодирования требуется столько бинарных признаков, сколько значений мог принимать исходный категориальный признак.

13. Рассмотрим решающее дерево для решения задачи кредитного скоринга, которое основано на двух признаках: возрасте и месячном доходе.



Пусть дана следующая выборка из пяти объектов (первый признак — возраст, второй — месячный доход):

[20, 8000], [15, 15000], [28, 9500], [24, 30000], [30, 20000]

Для сколько из них будет принято положительное решение о выдаче кредита?

2

14. Как в общем устроен процесс построения решающего дерева?

- ☐ Жадно — начинаем с одной вершины, разбиваем её на две, после чего рекурсивно повторяем процедуру для новых дочерних вершин.

Правильно

Верно.

- ☐ Жадно — начинаем с дерева, у которого в каждом листе находится по одному объекту, и удаляем из него вершины, пока улучшается качество.
- ☐ Полным перебором — вычисляем качество каждого возможного дерева, выбираем лучшее.
- ☐ Аналитически — можно в явном виде выписать формулы, задающие структуру оптимального дерева.

15. Мы пытаемся найти лучшее разбиение вершины  $m$ , и хотим оценить качество конкретного способа, который разобьёт вершину  $m$  на вершины  $l$  и  $r$ . Обозначим через  $|X_m|, |X_l|, |X_r|$  количество объектов в вершинах  $m, l$  и  $r$  соответственно, через  $H(X)$  - значение критерия информативности на выборке  $X$ .

По какой из формул следует вычислять ошибку такого разбиения?

- ☐  $\frac{|X_l|}{|X_m|} H(X_l) + \frac{|X_r|}{|X_m|} H(X_r)$
- ☒  $\frac{|X_l|}{|X_m|} \frac{H(X_l)}{H(X_m)} + \frac{|X_r|}{|X_m|} \frac{H(X_r)}{H(X_m)}$
- ☐  $\frac{H(X_l)}{H(X_m)} + \frac{H(X_r)}{H(X_m)}$
- ☐  $H(X_l) + H(X_r)$

16. Можно ли решать задачу регрессии с помощью решающих деревьев?

- ☐ Да, можно — достаточно лишь выбрать критерий информативности, оценивающий разброс вещественных ответов.

- ☐ Нет, нельзя — деревья могут выдавать столько различных ответов, сколько в дереве листьев, то есть конечное число. А в задаче регрессии бесконечно много возможных ответов.
- ☐ Нет, нельзя — критерии информативности зависят от распределения объектов по классам, такие распределения нельзя построить в задачах регрессии.

17. Выберите верные утверждения про обучение случайного леса.

- ☐ Каждое дерево обучается по случайной подвыборке признаков.
- ☐ Как правило, строятся деревья небольшой глубины, поскольку этого достаточно для восстановления сложных зависимостей.
- ☐ Каждое дерево обучается по случайной подвыборке объектов.
- ☐ Каждое дерево обучается независимо от остальных деревьев в композиции.
- ☐ В каждой вершине оптимальный признак для разбиения выбирается из случайного подмножества признаков.

18. Какие величины предсказывает  $N$ -й базовый алгоритм в градиентном бустинге?

- ☒ Производные (с минусом) функции потерь, вычисленные в точках, соответствующих ответам композиции  $a_{N-1}(x)$  на обучающей выборке.

Правильно

- ☐ Разницу между истинными ответами и ответами композиции  $a_{N-1}(x)$  на обучающей выборке.
- ☐ Производные (с минусом) функции потерь, вычисленные в точках, соответствующих ответам предыдущего базового алгоритма  $b_{N-1}(x)$  на обучающей выборке.

19. Чем градиентный бустинг отличается от случайного леса?

- ☐ Базовые алгоритмы, как правило, выбираются достаточно простыми — например, это могут быть неглубокие деревья.
- ☐ Градиентный бустинг может строить алгоритмы только для задач регрессии.
- ☐ Каждый следующий алгоритм в градиентном бустинге обучается так, чтобы исправить ошибки предыдущих базовых алгоритмов.

20. В задаче кластеризации мы пытаемся:

- ☐ Восстановить отображение вектора признаков в метку класса по набору известных пар (признаки, метка).
- ☐ Восстановить отображение вектора признаков в действительное число по набору известных пар (признаки, метка).
- ☐ Сгруппировать похожие объекты (близкие в пространстве признаков) в кластеры, поместив при этом непохожие друг на друга объекты в разные кластеры.

21. Какие параметры есть у метода  $k$  ближайших соседей?

☐ Число деревьев

правильно, этот вариант не должен быть выбран

☒ Число соседей  $k$

Правильно

☐ Функция активации

правильно, этот вариант не должен быть выбран

☒ Функция весов объектов

Правильно

22. Какие из перечисленных алгоритмов нуждаются в задании количества кластеров?

☐ DBSCAN

☐ k-Means

☐ EM-алгоритм



23. С помощью какого алгоритма можно разделять смеси распределений?

- ☐ Алгоритм К-средних
- ☐ EM-алгоритм

Правильно

Верно.

- ☐ Алгоритм Прима

24. Выберите верные утверждения про метод главных компонент

- ☐ Метод главных компонент — это метод понижения размерности с потерями, если число новых признаков  $d$  меньше ранга исходной матрицы "объекты-признаки".
- ☐ Метод главных компонент является линейным методом понижения размерности — новые признаки являются линейными комбинациями исходных признаков.
- ☐ Метод главных компонент позволяет найти нелинейную поверхность, при проецировании выборки на которую дисперсия оказывается максимальной.
- ☐ Матрица весов  $W$ , которая строится в методе главных компонент, является ортогональной.

25. Что неотрицательного в методе неотрицательных матричных разложений?

- ☒ Элементы раскладываемой матрицы

Правильно

Потому что элементы приближающей матрицы неотрицательны и было бы странно приближать такой матрицей, матрицу с отрицательными элементами.

- ☐ Определители всех угловых миноров раскладываемой матрицы

правильно, этот вариант не должен быть выбран

☐ Определители матриц-множителей

правильно, этот вариант не должен быть выбран

☒ Элементы матриц-множителей

Правильно

Это верно по определению неотрицательного матричного разложения.

☒ Значения функции потерь

Правильно

Как и во всех остальных задачах, функция потерь отрицательных значений не принимает.

26. Какую функцию активации  $\sigma$  нужно взять в однослойной нейронной сети  $a(x, w) = \sigma(\sum_{j=1}^d w_j x_j + w_0)$ , чтобы получилась линейная регрессия?

☐  $\sigma(u) = u$

Правильно

☐  $\sigma(u) = \text{sign}(u)$

☐  $\sigma(u) = \frac{1}{1+e^{-u}}$

27. Какую функцию активации  $\sigma$  нужно взять в однослойной нейронной сети  $a(x, w) = \sigma(\sum_{j=1}^d w_j x_j + w_0)$ , чтобы получилась классификация (не логическая регрессия) ?

☐  $\sigma(u) = u$

☒  $\sigma(u) = \text{sign}(u)$

Правильно

☐  $\sigma(u) = \frac{1}{1+e^{-u}}$

28. Какие параметры являются мерами разброса распределений?

☐ Математическое ожидание.

правильно, этот вариант не должен быть выбран

☒ Среднеквадратическое отклонение.

Правильно

Среднеквадратическое отклонение — это корень из дисперсии; как и дисперсия, оно характеризует разброс распределения.

☒ Интерквартильный размах.

Правильно

Интерквартильный размах — мера разброса.

☐ Медиана.

правильно, этот вариант не должен быть выбран

☒ Дисперсия.

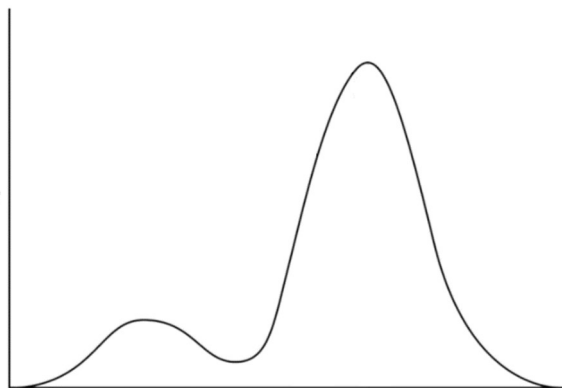
Правильно

Дисперсия — наиболее часто используемая мера разброса.

☐ Мода.

правильно, этот вариант не должен быть выбран

29. Пусть график плотности распределения случайной величины  $X$  выглядит следующим образом:



Выберите верные утверждения о средних такой случайной величины.

☐ Медиана  $X$  меньше её математического ожидания

правильно, этот вариант не должен быть выбран

Медиана менее чувствительна к выбросам, в свою очередь небольшое количество экстремальных значений сильно влияют на математическое ожидание.

☒ Мода  $X$  больше её медианы.

Правильно

Мода совпадает с положением большого пика, а медиана смещается в сторону меньшего.

☐ Математическое ожидание и медиана  $X$  совпадают, а мода не определена.

правильно, этот вариант не должен быть выбран

☒ Мода  $X$  больше её математического ожидания.

Правильно

Мода совпадает с положением большого пика, а математическое ожидание смещается в сторону меньшего.

30. Пусть  $X \sim F(x)$  — случайная величина с произвольным распределением, неизвестным математическим ожиданием  $EX$  и известной дисперсией  $DX$ . Как выглядит доверительный интервал для  $EX$  с приближённым уровнем доверия  $100(1 - \alpha) \%$ ,  $N$  – объем выборки?

☒  $\bar{X} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{DX}{N}}$

Правильно

☐  $\bar{X} \pm z_{1-\frac{\alpha}{2}} \frac{DX}{\sqrt{N}}$

☐  $\bar{X} \pm t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{DX}{N}}$

☐  $\bar{X} \pm z_{1-\alpha} \sqrt{\frac{DX}{N}}$

31. Пусть  $H_0$  - нулевая, основная гипотеза,  $H_1$  – альтернативная, конкурирующая гипотеза, если достигаемый уровень значимости  $p \leq \alpha$ , то:

☐  $H_0$  не отвергается

☐  $H_1$  верна

☐  $H_0$  отвергается в пользу  $H_1$

Правильно

☐  $H_1$  отвергается в пользу  $H_0$

☐  $H_0$  верна

32. При проверке некоторой гипотезы значение статистики составило 54123432,22. Достигаемый уровень значимости  $p$  — это:

☐ вероятность справедливости нулевой гипотезы.

☐ вероятность справедливости нулевой гипотезы при таком значении статистики.

☐ вероятность получить такое значение статистики.

☒ ничего из перечисленного

Правильно

Достижимый уровень значимости — вероятность получить такое или ещё более экстремальное значение статистики при справедливости нулевой гипотезы.

33. Ошибка первого рода — это:

☒ отвержение верной нулевой гипотезы

Правильно

☐ принятие неверной нулевой гипотезы

34. Объём выборки, необходимый для построения доверительного интервала заданной ширины:

☒ увеличивается с уменьшением требуемой ширины

Правильно

☐ увеличивается с увеличением требуемой ширины

правильно, этот вариант не должен быть выбран

☒ увеличивается с ростом дисперсии выборки

Правильно

☐ увеличивается с ростом  $\alpha$

правильно, этот вариант не должен быть выбран

☒ уменьшается с ростом  $\alpha$

Правильно

☐ уменьшается с ростом дисперсии выборки

правильно, этот вариант не должен быть выбран

35. С помощью каких из этих величин по данным социологического опроса можно оценить силу взаимосвязи между возрастом (в годах) и средним годовым доходом (в рублях)?

☐ Кoeffициент V Крамера

правильно, этот вариант не должен быть выбран

☒ Кoeffициент корреляции Пирсона

Правильно

Так мы измерим силу линейной взаимосвязи между нашими непрерывными признаками

☒ Кoeffициент корреляции Спирмена

Правильно

Так мы измерим силу монотонной взаимосвязи между нашими непрерывными признаками

☐ Разность между средними доходами молодых и старых

правильно, этот вариант не должен быть выбран

☐ Кoeffициент корреляции Мэтьюса

правильно, этот вариант не должен быть выбран

36. С помощью каких из этих величин по данным социологического опроса можно оценить силу взаимосвязи между полом и семейным положением (холост/в браке/в разводе)?

☐ Разность между средними долями женатых среди мужчин и среди женщин

правильно, этот вариант не должен быть выбран

☒ Кoeffициент V Крамера

Правильно

☐ Кoeffициент корреляции Мэтьюса

правильно, этот вариант не должен быть выбран

☐ Коэффициент корреляции Спирмена

правильно, этот вариант не должен быть выбран

☐ Коэффициент корреляции Пирсона

правильно, этот вариант не должен быть выбран

37. Какой из статистических тестов используется для проверки на нормальность?

☐ Манна-Уитни

☒ Шапиро-Уилка

Правильно

☐ Стьюдента

☐ Вавилова-Черенкова

38. Какой из трёх методов поправки на множественную проверку позволяет отвергнуть больше всего гипотез?

☐ Бенджамини-Хохберга

Правильно

☐ Холма

☐ Бонферрони