

# Тест для аналитиков

Ответьте, пожалуйста, на нижеприведенные вопросы настолько подробно насколько сможете, стараясь уложиться в указанный в вопросе временной интервал. На весь тест вам отводится 30 минут, так что время имеет значение.

Стремитесь давать ответы на языке математики. Идеальный ответ преимущественно состоит из формул и содержит минимум текста. Если можете дать ответ одним словом - так и сделайте.

Понятность и краткость в данном случае важнее математической строгости и полноты.

Если вы знаете больше, а места и времени не хватает, просто приведите перечень всех ключевых аспектов, имеющих отношение к теме - мы вас более подробно расспросим об этом после теста.

Нет необходимости отвечать на все вопросы. Выберите те, на которые вы можете дать быстрый и правильный ответ.

1. Дайте определение матрицы (**30 сек**)

2. Опишите модель линейной регрессии (**5 мин**)

3. Чем логистическая регрессия отличается от линейной (**5 мин**)

4. Возьмите производную от логистической функции (сигмоиды) (**1 мин**)
5. Выведите эту производную как функцию от самой сигмоиды (**2 мин**)
6. Выведите гиперболический тангенс как функцию от сигмоиды (**2 мин**)
7. Напишите формулу кросс энтропии для дискретного распределения (**15 сек**)
8. Выведите ее связь с расстоянием Кульбака-Лейблера? (**30 сек**)
9. Продифференцируйте для случая многоклассовой классификации кросс-энтропию по входным данным  $W$ , когда результат классификации  $\hat{y} = \text{softmax}(W)$  (**3 мин**)
10. Докажите что  $\text{softmax}$  инвариантен к смещению, т.е.  $\text{softmax}(x) = \text{softmax}(x + c)$  (**1 мин**)

11. Перечислите алгоритмы построения дерева решений (**30 сек**)

12. Опишите один из этих алгоритмов (**3 мин**)

13. Зачем логарифмировать функцию правдоподобия? (**1 мин**)

1.

2.

3.

4.

5.

14. Опишите метод градиентного спуска (**3 мин**)

15. Чем стохастический градиентный спуск отличается от градиентного спуска? (**30 сек**)

16. Перечислите методы регуляризации (**15 сек**)

17. Опишите один из них: в чем заключается, как работает, особенности, сложности, ... (**2 мин**)

18. Опишите еще один (**2 мин**)

19. Перечислите методы кластеризации (**15 сек**)

20. Опишите один из них (**3 мин**)