

## ПРИМЕРЫ ЭКЗАМЕНАЦИОННЫХ ВОПРОСОВ

по магистерской программе «Управление ИИ-продуктами/AI Product»  
Направление подготовки 02.04.03 Математическое обеспечение и  
администрирование информационных систем

### Алгоритмы и структуры данных

1. Что означает следующая запись?  $g(x) = o(f(x))$
2. Сколькими способами мы можем выбрать различные подмножества из множества мощности  $n$ ?
3. Какие знаете алгоритмы, в которых средняя амортизированная сложность операции имеет асимптотику ниже реальной.
4. Какая сложность у алгоритма быстрой сортировки?

### Программирование на языке Python

5. Что вернет выражение  $s * n$ , где  $s$  - строка,  $n$  - целое число
6. В чем разница между generator и list comprehension?
7. Что может быть ключем у dict?
8. Что такое инкапсуляция?
9. Какие знаете особенности реализации ООП на python?
10. Зачем нужны метаклассы в python?
11. Что такое ленивые вычисления в python, как реализованы?

### DevOps и MLOps

12. Как вывести имя текущей командной оболочки?
13. Какие действия могут привести к удалению Git-коммита из локального репозитория?
14. В чем ключевая разница между DVC и LakeFS?

15. Какая функциональность есть в MLFlow?

### **Базы данных. SQL и NoSQL**

16. В чем разница между SQL и NoSQL?

17. В чем разница между LEFT JOIN, RIGHT JOIN, INNER JOIN, UNION?

18. Выберите верные утверждения про MongoDB и MySQL

### **Основные концепции Big Data**

19. Какие из следующих распределенных СУБД не относятся к NoSQL?

20. Как происходит удаление данных в HDFS?

21. Что из перечисленного отвечает за распределение ресурсов  
высокопараллельному приложению?

22. Какие копии создаются при репликации в HDFS?

### **Теория вероятностей и математическая статистика**

23. На овощном рынке продаются биткойны. 80% биткойнов приехали из Азии, а оставшиеся 20% — из Европы. 20% азиатских и 5% европейских биткойнов — фальшивые. Чему равна вероятность купить фальшивый биткойн на этом рынке? Ответ округлите до 2 знака после запятой

24. С помощью каких функций можно задавать распределения непрерывных случайных величин?

25. Пусть  $x$  — непрерывная случайная величина, задана своей плотностью. К какому значению стремится вероятность  $P(x = m)$ ?

26. Какие значения принимает функция распределения  $F(x)$  случайной величины в пределе при  $x$ , стремящемся к  $-\infty$  и стремящемся к  $+\infty$

## **Основы машинного обучения**

27. Чем отличается задача регрессии от задачи классификации?

28. У вас есть датасет из 500 размеченных сэмплов. Вы обучили модель и, в первом случае, проводите k-fold кросс валидацию на 500 фолдов. Во втором случае вы имеете ту же самую модель и используете leave-one-out кросс валидацию. Какой способ, вероятнее всего, покажет худшие метрики?

29. Чему будет равен ROC-AUC константного ответа 0 на выборке из 99000 объектов класса 0 и 1000 объектов класса 1.

30. Выберите верные утверждения про градиентный бустинг над решающими деревьями и случайный лес.

## **Нейронные сети**

31. Какие методы оптимизации используются для обучения нейронной сети?

32. Соотнесите графики функций активации и их названия

(1 - sigmoid, 2 - ReLU, 3 - ELU, 4 - tanh; графики от а) до г) слева направо сверху вниз):

33. Во время обучения нейронной сети выяснилось, что на начальных эпохах функция потерь не падала. С чем это связано?

## Основы компьютерного зрения (CV)

- 34. Предположим на вход сверточному слою сети с  $K$  фильтрами размера 5 подается RGB картинка размера  $512 \times 256$ . Сколько обучаемых параметров получится в таком слое?
- 35. Какие архитектуры традиционно относятся к задачам детекции (segmentation, object detection)?
- 36. Каким образом можно повысить variance в задаче классификации картинок с помощью сверточных сетей?

## Основы обработки естественного языка (NLP)

- 37. В чем состоит задача Named Entity Recognition?
- 38. Что такое NLTK?
- 39. Accuracy в NLP
- 40. Опишите архитектуру transformer.

## Задачи на коддинг:

- 1. Вам дан датасет, посчитайте статистические характеристики - среднее, медиану, среднеквадратичное отклонение по выбранным признакам, отфильтруйте датасет по полю X.
- 2. Реализуйте приложение с LLM на python с использованием любого доступного вам инструментария и приложите ссылку на репозиторий с кодом
- 3. Проведите разведочный анализ данных для предложенного датасета

## Soft skills

Что вы сделаете в предложенной ситуации? (даны кейсы-ситуации)

- Вы заметили, что коллега уже третий раз делает одну и ту же неточность в коде, из-за чего немного падает производительность сервиса. На еженедельном статус-звонке два архитектора спорят о тонкостях алгоритма, остальные молчат, время заканчивается.
- За два дня до релиза клиент просит «маленькую» дополнительную фичу, не учтённую в спринте.
- Вы поняли, что дали менеджеру заниженный estimate: реальная работа займёт на два дня больше.
- Вас просят выступить с часовым воркшопом по продукту, в который у вас неглубокое погружение.