

1. Что такое данные в контексте машинного обучения? Приведите примеры, чем отличаются категориальные признаки от числовых?
2. Как можно преобразовать категориальные признаки в числовые? Объясните методы.
3. В чем заключается основная идея линейной регрессии? Что представляет собой модель линейной регрессии? Напишите уравнение.
4. Что такое функция потерь и чем отличается от функционала ошибки? Напишите формулу для MSE и MAE.
5. Опишите процесс обучения модели линейной регрессии.
6. Какова основная цель метода наименьших квадратов? Напишите ход работы этого метода.
7. Что такое коэффициент детерминации (R^2)? Как его интерпретировать?
8. Напишите формализацию градиентного спуска и ход работы этого метода.
9. Чем лучше стохастический градиентный спуск в отличие от обычного градиентного спуска? Напишите 3 варианта, чтобы остановить алгоритм градиентного спуска.
10. Почему масштабирование признаков важно и как влияет на работу градиентного спуска?
11. Какие способы нормализации данных вы знаете?
12. Что такое MinMax-масштабирование, напишите формулу и также реализуйте на языке Python.
13. Объясните, в чем отличие полиномиальной регрессии от линейной регрессии? Какую роль играет степень полинома, и как она влияет на качество модели и ее способность к обобщению?
14. Что такое переобучение в контексте линейной регрессии и какие признаки указывают на переобучение?
15. Что такое регуляризация? Объясните, что такое L1-регуляризация с формулами и нарисуйте график линий уровней L1-регуляризации. Что такое коэффициент регуляризации и какую роль играет?
16. Что такое регуляризация? Объясните, что такое L2-регуляризация с формулами и нарисуйте график линий уровней L2-регуляризации. Что такое коэффициент регуляризации и какую роль играет?
17. Что такое регуляризация? Объясните, что такое L1+L2-регуляризация с формулами и нарисуйте график линий уровней L1+L2-регуляризации. Что такое коэффициент регуляризации и какую роль играет?

18. Что такое функция правдоподобия в логистической регрессии? Напишите формулу.
19. Какие функции потерь можно использовать для задачи классификации? Почему MSE не подходит?
20. Что такое нормальные уравнения? Напишите формулу. У метода нормальных уравнений есть несколько проблем, напишите какие.
21. Что представляет собой метод максимального правдоподобия? Как метод максимального правдоподобия применяется при обучении моделей классификации?
22. Что такое сигмоидная функция? Напишите её формулу. Какова роль сигмоида в логистической регрессии?
23. Как значения, полученные из сигмоида, интерпретируются в задачах классификации?
24. Что такое распределение Бернулли? Приведите его вероятностную функцию. Как распределение Бернулли связано с задачами бинарной классификации?
25. Чем логистическая регрессия отличается от линейной? Какова цель логистической регрессии? Напишите её основное уравнение.
26. Какие метрики качества используются для оценки моделей классификации?
27. Что такое матрица ошибок (confusion matrix)? Объясните её компоненты.
28. Определите accuracy, recall, precision и F1-меру. Когда какую метрику лучше использовать?