1. Задача обучения с учителем
2. Задача обучения без учителя. Предоставьте по крайней мере две проблемы-примера.
3. Что такое i.i.d. данные?
4. Как работает наивный байесовский классификатор? Почему он наивный?
5. Модель линейной регрессии для минимизации проблемы MSE. Запишите формулу и производную функции потерь по весам.
6. Запишите шаг градиентного спуска. Как его адаптировать для больших наборов данных?
7. Что такое правдоподобие? Где обычно используется Максимальное Правдоподобие (MLE)?
8. Что такое кросс-валидация? Как количество фолдов влияет на валидацию?
9. Что такое переобучение и недообучение? Как их обнаружить?
10. В чем разница между параметрами и гиперпараметрами? Приведите пример для линейных моделей/деревьев решений.
11. Что такое регуляризация? В чем разница между L1 и L2 регуляризацией в линейных моделях? Это единственный способ ограничить решение?
12. Регуляризует ли L2 регуляризация член смещения (w\_0 или b)? Почему?
13. Почему хорошая идея нормализовать данные перед применением линейной модели?
14. Сформулируйте задачу линейной классификации. Что такое маржа?
15. Что такое точность и полнота? Как их использовать для оценки качества модели?
16. Предположим, что набор данных для бинарной классификации несбалансирован, так что 95% данных относятся к первому классу. Как скорректировать меры качества классификации для работы с такими данными? Почему?
17. Что такое ROC AUC? Как построить ROC-кривую?
18. Функция потерь логистической регрессии. Как она связана с оценкой максимального правдоподобия?
19. Основная идея машины опорных векторов. Функционал оптимизации для линейно разделимого случая.
20. Опишите жадный оптимизационный алгоритм для дерева решений.
21. Почему неограниченное дерево решений может достичь нулевой ошибки на обучающем наборе с уникальными объектами?
22. Как присвоить метку класса объекту в листе дерева в классификации?
23. Как присвоить метку класса объекту в листе дерева в регрессии? Зависит ли это от информационного критерия?
24. Что такое бэггинг?
25. Что такое Случайный лес? Чем он отличается от Бэггинга над деревьями решений?
26. Как обучаются базовые алгоритмы в градиентном бустинге?
27. Как работает обратное распространение ошибок в нейронных сетях? Что будет вектор-векторная производная?
28. Как работает свёрточный слой? Что такое свёрточная операция?
29. Почему полносвязные (плотные) сети не лучший выбор для работы с изображениями? Почему лучше работают свёрточные нейронные сети (CNN)?
30. Как работает базовая RNN (Vanilla RNN)?
31. Как работает Dropout?
32. Как Dropout и пакетная нормализация меняют свое поведение на этапе вывода?
33. Сформулируйте оптимизационную задачу для Метода Главных Компонент?