1. Понятие машинного обучения. Отличие машинного обучения от других областей программирования.
2. Классификация задач машинного обучения. Примеры задач из различных классов.
3. Основные понятия машинного обучения: набора данных, объекты, признаки, атрибуты, модели, параметры.
4. Структура и представление данных для машинного обучения.
5. Инструментальные средства машинного обучения.
6. Задача регрессии: постановка, математическая формализация.
7. Метод градиентного спуска для парной линейной регрессии.
8. Понятие функции ошибки: требования, использование, примеры.
9. Множественная и нелинейная регрессии.
10. Нормализация признаков в задачах регрессии.
11. Задача классификации: постановка, математическая формализация.
12. Метод градиентного спуска для задач классификации.
13. Логистическая регрессия в задачах классификации.
14. Множественная и многоклассовая классификация. Алгоритм “один против всех”.
15. Метод опорных векторов в задачах классификации.
16. Понятие ядра и виды ядер в методе опорных векторов.
17. Метод k ближайших соседей в задачах классификации.
18. Метод решающих деревьев в задачах классификации.
19. Однослойный перцептрон в задачах классификации.
20. Метрики эффективности и функции ошибки: назначение, примеры, различия.
21. Предварительный анализ данных: задачи, методы, цели.
22. Понятие набора данных (датасета) в машинном обучении. Требования, представление. Признаки и объекты.
23. Шкалы измерения признаков. Виды шкал, их характеристика.
24. Понятие чистых данных. Определение, очистка данных.
25. Основные этапы проекта по машинному обучению.
26. Проблема отсутствующих данных: причины, исследование, пути решения.
27. Проблема несбалансированных классов: исследование, пути решения.
28. Понятие параметров и гиперпараметров модели. Обучение параметров и гиперпараметров. Поиск по сетке.
29. Понятие недо- и переобучения. Определение, пути решения.
30. Диагностика модели машинного обучения. Методы, цели.
31. Проблема выбора модели машинного обучения. Сравнение моделей.
32. Измерение эффективности работы моделей машинного обучения. Метрики эффективности.
33. Метрики эффективности моделей классификации. Виды, характеристика, выбор.
34. Метрики эффективности моделей регрессии. Виды, характеристика, выбор.
35. Перекрестная проверка (кросс-валидация). Назначение, схема работы.
36. Конвейеры в библиотеке sklearn. Назначение, использование.
37. Использование методов визуализации данных для предварительного анализа.
38. Исследование коррелированности признаков: методы, цели, выводы.
39. Решкалирование данных. Виды, назначение, применение. Нормализация и стандартизация данных.
40. Преобразование категориальных признаков в числовые.
41. Проблема сбора и интеграции данных для машинного обучения.
42. Понятие чистых данных и требования к данным.
43. Основные задачи описательного анализа данных.
44. Методы визуализации данных для машинного обучения.
45. Задача выбора модели. Оценка эффективности, валидационный набор.
46. Кривые обучения для диагностики моделей машинного обучения.
47. Регуляризация моделей машинного обучения. Назначение, виды, формализация.
48. Полиномиальные модели машинного обучения.
49. Основные виды преобразования данных для подготовки к машинному обучению.
50. Задача выбора признаков в машинном обучении.