

Вопросы к экзамену по ML 2018

Основные определения

- 1 Основные определения в машинном обучении: объект, целевая функция, признак, модель, обучающая выборка, функционал качества, обучение, переобучение.
- 2 Задачи машинного обучения - обучение с учителем, без учителя. Задачи регрессии и классификации. Задачи снижения размерности и кластеризации.
- 3 Типы признаков в машинном обучении. Приведите примеры различных признаков.

Метрические классификаторы

- 4 Метод k ближайших соседей в задаче классификации.
- 5 Методы отбора признаков. Жадный метод.
- 6 Определение отступа в метрических алгоритмах классификации. Алгоритм Condensed Nearest Neighbor.
- 7 Метод k ближайших соседей в задаче регрессии.
- 8 Обобщение метода k ближайших соседей через взвешенный учет объектов. Ядерная оценка плотности.
- 9 Проклятие размерности. Зависимость метода ближайших соседей от масштабирования признаков. Способы стандартизации признаков.

Кластеризация

- 10 Постановка задачи кластеризации. Цели кластеризации. Типы кластерных структур. Чувствительность к нормировке и масштабированию признаков.
- 11 Метод k средних. Особенности метода.

12 Степени свободы метода k средних. Метод k -means++. Метод Xmeans.

13 Графовые алгоритмы кластеризации.

14 Алгоритм Ланса-Уильямса.

Деревья принятия решений

15 Логическая закономерность. Интерпретируемость и информативность.

16 Решающий список. Достоинства и недостатки.

17 Структура решающего дерева, метод спуска по дереву в общем случае. Достоинства и недостатки решающих деревьев.

18 Подрезание решающих деревьев.

19 Небрежные решающие деревья.

20 Деревья принятия решений в задаче регрессии.

Байесовские методы

21 Вероятностная постановка задачи классификации. Функция правдоподобия и априорная вероятность.

22 Функционал среднего риска. Общая формула байесовского классификатора.

23 Наивный байесовский классификатор.

24 Восстановление плотности распределения по выборке.

25 Аддитивное сглаживание для байесова классификатора.

Линейные классификаторы

26 Модель МакКаллока-Питтса

27 Обобщённая модель линейного классификатора. Определение отступа. Минимизация эмпирического риска.

28 Метод градиентного спуска. Выбор величины шага.

29 L_2 регуляризация.

30 Метод стохастического градиента. Недостатки метода SG и как с ними бороться.

Способность к обобщению

31 Внутренний и внешний функционал качества. Кросс-валидация.

32 Критерий непротиворечивости моделей.

33 Аналитическая оценка вероятности переобучения. Схема использования.

34 Неравенство Бернштейна-Хёфдинга в применении к задаче выбора модели.

35 Дихотомии. Функция роста. Точка разрыва.

36 Оценка на максимальное число дихотомий.

Нейронные сети

37 Представимость функций в виде нейросети.

38 Метод обратного распространения ошибок. Основные недостатки и способы их устранения.

39 Выбор начального приближения в градиентных методах настройки нейронных сетей. Функции активации.

40 Устройство свёрточной нейросети.

41 Нейронные сети для задачи регрессии.

Метод опорных векторов

- 42** Постановка задачи SVM.
- 43** Регуляризация в задаче SVM.
- 44** Двойственная задача SVM.
- 45** Ядерный алгоритм SVM.
- 46** Представление метода опорных векторов в виде нейронной сети.
- 47** Метод SVR для задачи регрессии.

Линейная регрессия

- 48** Постановка задачи многомерной линейной регрессии. Матричная запись.
- 49** Использование сингулярного разложения для решения задачи наименьших квадратов.
- 50** Проблема «мультиколлинеарности» в задачах многомерной линейной регрессии.
- 51** Гребневая регрессия. Регуляризация Лассо.
- 52** Нелинейная регрессия. Метод Ньютона-Гаусса.
- 53** Задача уменьшения размерности. Метод главных компонент.

Анализ смещения и разброса

- 54** Постановка задачи анализа смещения и разброса. Качество обучения в зависимости от пространства моделей.
- 55** Внутренний и внешний функционал качества. Средний метод.
- 56** Определение смещения и разброса.

57 Кривые обучения.

Ансамбли

58 Определение композиции алгоритмов. Типы композиций.

59 Взвешенное голосование. Бустинг. Алгоритм AdaBoost.

60 Простое голосование. Бэггинг и метод случайных подпространств.

61 Случайный лес. Стэкинг.