

Теоретический зачет
Параллель А-МЛ

ЛКШ.2017.Август

1. Что такое машинное обучение (МО)? Приведите пример задачи, которую можно решить с помощью МО. Что такое объект и его признаковое описание?
2. Перечислите основные типы задач МО.
3. Что такое обучающая и тестовая выборки? Что такое алгоритм (модель)? Как можно измерить качество работы алгоритма?
4. Что такое кросс-валидация и зачем она нужна? Расскажите алгоритм кросс-валидирования. Приведите пример кросс-валидации на логистической регрессии.
5. Что такое функция потерь? Как она связана с функционалом качества (ошибки)? Приведите примеры функций потерь.
6. Какие виды предобработки данных вы знаете?
7. Что такое переобучение? Как с ним бороться?
8. Перечислите типы признаков и способы работы с категориальными признаками.
9. Что такое регуляризация? Какие виды регуляризации вы знаете? Зачем она нужна? Приведите пример модели с l_2 -регуляризацией. Какая регуляризация довольно устойчива к выбросам?
10. Расскажите про линейную и логистическую регрессии.
11. Какие способы предобработки текстов вы знаете?
12. Какие способы выделения признаков из текста вы знаете?
13. Как работает алгоритм kNN? Какие типы метрик вы знаете?
14. Напишите формулу предсказания алгоритма kNN в случае задачи классификации и регрессии. Почему kNN чувствителен к масштабированию признаков?
15. Как работает решающее дерево?
16. Какие критерии останова построения решающего дерева вы знаете? Устойчиво ли решающее дерево к нормированию признаков? Как строятся предсказания в случае задачи классификации и в случае задачи регрессии.

17. Как работает Random Forest (RF)? Какой глубины нужны деревья для построения ансамбля? Переобучается ли RF с ростом количества обучаемых деревьев? Как работает алгоритм Extremely Randomized Trees?
18. Архитектура многослойного персептрона (нейронная сеть). Какие функции активации вы знаете? Интуиция алгоритма обратного распространения ошибки.
19. Операция свертки над изображением. Ядро свертки. Перечислите основные фильтры и что они делают
20. Архитектура сверточной нейронной сети
21. Способы понижения размерности пространства. Чем они отличаются?
22. Алгоритмы кластеризации: kMeans, Mini-batch kMeans, Иерархические кластеризации: bottom-up и top-down.
23. Алгоритмы кластеризации: DBSCAN, Grid-based. Оценка качества кластеризации.