

Вопросы к контрольной работе
Машинное обучение 1
ПМИ ФКН ВШЭ

1. Что такое объект, целевая переменная, признак, модель, функционал ошибки и обучение?
2. Запишите формулы для линейной модели регрессии и для среднеквадратичной ошибки. Запишите среднеквадратичную ошибку в матричном виде.
3. Что такое коэффициент детерминации? Как интерпретировать его значения?
4. Чем отличаются функционалы MSE и MAE?
5. Как устроены робастные функции потерь (Huber loss, log-cosh)? Чем log-cosh лучше функции потерь Хубера?
6. Что такое градиент? Какое его свойство используется при минимизации функций?
7. Как устроен градиентный спуск?
8. Почему не всегда можно использовать полный градиентный спуск? Какие способы оценивания градиента вы знаете? Почему в стохастическом градиентном спуске важно менять длину шага по мере итераций? Какие стратегии изменения шага вы знаете?
9. В чём заключаются метод инерции и AdaGrad/RMSProp?
10. Что такое кросс-валидация? На что влияет количество блоков в кросс-валидации? Как построить итоговую модель после того, как по кросс-валидации подобраны оптимальные гиперпараметры?
11. Чем гиперпараметры отличаются от параметров? Что является параметрами и гиперпараметрами в линейных моделях и в решающих деревьях?
12. Что такое регуляризация? Запишите L1- и L2-регуляризаторы.
13. Почему L1-регуляризация отбирает признаки?
14. Почему плохо накладывать регуляризацию на свободный коэффициент? Почему регуляризация может странно работать на немасштабированных данных?
15. Запишите формулу для линейной модели классификации. Что такое отступ? Как обучаются линейные классификаторы и для чего нужны верхние оценки пороговой функции потерь?
16. Что такое точность, полнота и F-мера? Почему F-мера лучше арифметического среднего и минимума?
17. Для чего нужен порог в линейном классификаторе? Из каких соображений он может выбираться?
18. Что такое AUC-ROC? Опишите алгоритм построения ROC-кривой.
19. Что такое AUC-PRC? Опишите алгоритм построения PR-кривой.
20. Что означает “модель оценивает вероятность положительного класса”? Как можно внедрить это требование в процедуру обучения модели?
21. Что такое калибровочная кривая? Какие методы калибровки вероятности вы знаете? Почему важно проводить калибровку не на обучающей выборке?

22. Запишите функционал логистической регрессии. Как он связан с методом максимума правдоподобия?
23. Запишите задачу метода опорных векторов для линейно неразделимого случая. Как функционал этой задачи связан с отступом классификатора? Как выглядит задача безусловной оптимизации в SVM?
24. В чём заключаются one-vs-all и all-vs-all подходы в многоклассовой классификации?
25. Как измеряется качество в задаче многоклассовой классификации? Что такое микро- и макро-усреднение?
26. В чём заключается преобразование категориальных признаков в вещественные с помощью mean-target encoding? Почему использование этого способа кодирования может привести к переобучению? Какие методы борьбы с этой проблемой вам известны?
27. Как устроены методы bag-of-words и tf-idf для кодирования текстов?
28. Как определить для линейной модели, какие признаки являются самыми важными?
29. Опишите жадный алгоритм обучения решающего дерева.
30. Почему с помощью бинарного решающего дерева можно достичь нулевой ошибки на обучающей выборке без повторяющихся объектов?
31. Как в общем случае выглядит критерий хаотичности? Как он используется для выбора предиката во внутренней вершине решающего дерева? Как вывести критерий Джини и энтропийный критерий?
32. Для какой ошибки строится разложение на шум, смещение и разброс? Запишите формулу этой ошибки.
33. Запишите формулы для шума, смещения и разброса метода обучения для случая квадратичной функции потерь.
34. Приведите пример семейства алгоритмов с низким смещением и большим разбросом; семейства алгоритмов с большим смещением и низким разбросом. Поясните примеры.
35. Что такое бэггинг? Как его смещение и разброс связаны со смещением и разбросом базовых моделей?
36. Что такое случайный лес? Чем он отличается от бэггинга над решающими деревьями?
37. Что такое out-of-bag оценка в бэггинге?
38. Запишите вид композиции, которая обучается в градиентном бустинге. Как выбирают количество базовых алгоритмов в ней?
39. Что такое сдвиги в градиентном бустинге? Как они вычисляются и для чего используются?
40. Как обучается очередной базовый алгоритм в градиентном бустинге? Что такое сокращение шага?
41. В чём заключается переподбор прогнозов в листьях решающих деревьев в градиентном бустинге?
42. Как в xgboost выводится функционал ошибки с помощью разложения в ряд Тейлора?
43. Какие регуляризации используются в xgboost?
44. Задача кластеризации. Метрики качества.

45. Метод K-Means, вывод его шагов.

46. Метод DBSCAN.

47. Метод t-SNE: основные формулы и функционал ошибки.