

**Билеты для
коллоквиума по
машинному
обучению ФЭН 2020**

1. Типы задач в машинном обучении (с примерами), обучение с учителем и без учителя, обучение модели, оценка качества модели, алгоритм решения задачи анализа данных.
2. Линейная модель регрессии. Аналитическое решение для среднеквадратичной ошибки (с выводом). Связь метода максимального правдоподобия с методом наименьших квадратов.
3. Линейная регрессия. Градиентный спуск. Градиентное обучение линейной регрессии.
4. Стохастический градиентный спуск, его модификации.
5. Функционалы ошибки для регрессии: MSE, MAE, коэффициент детерминации, квантильные потери, MAPE, SMAPE. В каких случаях какими из них стоит пользоваться.
6. Регуляризация. Аналитический вид вектора весов в линейной регрессии со среднеквадратичной ошибкой и квадратичным регуляризатором (с выводом).
7. Разреженные модели и L1-регуляризация. Почему использование L1-регуляризатора приводит к отбору признаков?
8. Линейная модель классификации. Отступ. Обучение линейных классификаторов через аппроксимацию функции $\text{sign}(\cdot)$. Примеры таких функций.
9. Функционалы ошибки для классификации: матрица ошибок, accuracy, precision, recall, F-мера.
10. Интегральные метрики. ROC-кривая и AUC-ROC, индекс Джини. Precision-recall-кривая и AUC-PR. Пример построения ROC-кривой и PR-кривой и вычисления AUC-ROC, AUC-PR.
11. Персептрон. Реализация логических функций (И, ИЛИ, НЕ) с помощью персептрона.
12. Логистическая регрессия. Оценивание вероятностей. Вывод логистической функции потерь из метода максимального правдоподобия.
13. Метод опорных векторов. Вывод постановки задачи для разделимого случая и для неразделимого случая.
14. Многоклассовая классификация: one-vs-all, all-vs-all. Какие проблемы возникают при использовании различных подходов? Метрики качества многоклассовой классификации.
15. Многоклассовая логистическая регрессия. Вывод многоклассового лог-лосса из метода максимального правдоподобия.
16. Работа с категориальными признаками: бинарное кодирование, хэширование, счётчики. Проблема переобучения счётчиков и борьба с ней.
17. Понятия ядра и спрямляющего пространства. Ядровой трюк. Теорема Мерсера. Примеры ядер, методы построения ядер.

Теоретический минимум

Незнание ответа на любой из нижеперечисленных вопросов влечёт за собой нулевую оценку за экзамен.

1. Что такое объект, целевая переменная, признак, модель, функционал ошибки и обучение?
2. Что такое переобучение и недообучение?
3. Задачи обучения с учителем и без учителя - определения и примеры.
4. Запишите формулы для линейной модели регрессии и для среднеквадратичной ошибки.
5. Что такое градиент? Какое его свойство используется при минимизации функций?
6. Запишите формулу для одного шага градиентного спуска. Какие способы оценивания градиента вы знаете? Зачем они нужны?
7. Что такое кросс-валидация? На что влияет количество блоков в кросс-валидации?
8. Что такое регуляризация? Почему L1-регуляризация отбирает признаки?
9. Запишите формулу для линейной модели классификации. Что такое отступ?
10. Что такое точность и полнота?
11. Что такое AUC-ROC? Как построить ROC-кривую?
12. Запишите функционал логистической регрессии. Как он связан с методом максимума правдоподобия?
13. Запишите задачу метода опорных векторов для линейно неразделимого случая. Как функционал этой задачи связан с отступом классификатора?
14. В чём заключаются one-vs-all и all-vs-all подходы в многоклассовой классификации?
15. В чём заключается преобразование категориальных признаков в вещественные с помощью счётчиков?
16. Что такое ядро?
17. Запишите определения полиномиального и гауссовского ядра.