# Билеты для экзамена по машинному обучению ФЭН 2020

1. Линейная модель регрессии. Аналитическое решение для среднеквадратичной ошибки (с выводом). Связь метода максимального правдоподобия с методом наименьших квадратов.
2. Линейная регрессия. Градиентный спуск. Градиентное обучение линейной регрессии.
3. Стохастический градиентный спуск, его модификации.
4. Функционалы ошибки для регрессии: MSE, MAE, коэффициент детерминации, квантильные потери, MAPE, SMAPE. В каких случаях какими из них стоит пользоваться.
5. Регуляризация. Аналитический вид вектора весов в линейной регрессии со среднеквадратичной ошибкой и квадратичным регуляризатором (с выводом).
6. Разреженные модели и L1-регуляризация. Почему использование L1-регуляризатора приводит к отбору признаков?
7. Линейная модель классификации. Отступ. Обучение линейных классификаторов через аппроксимацию функции sign(.). Примеры таких функций.
8. Функционалы ошибки для классификации: матрица ошибок, accuracy, precision, recall, F-мера.
9. 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.
10. Персептрон. Реализация логических функций (И, ИЛИ, НЕ) с помощью перспетрона.
11. Логистическая регрессия. Оценивание вероятностей. Вывод логистической функции потерь из метода максимального правдоподобия.
12. Метод опорных векторов. Вывод постановки задачи для разделимого случая и для неразделимого случая.
13. Многоклассовая классификация: one-vs-all, all-vs-all. Какие проблемы возникают при использовании различных подходов? Метрики качества многоклассовой классификации.
14. Многоклассовая логистическая регрессия. Вывод многоклассового лог-лосса из метода максимального правдоподобия.
15. Работа с категориальными признаками: бинарное кодирование, хэширование, счётчики. Проблема переобучения счётчиков и борьба с ней.
16. Понятия ядра и спрямляющего пространства. Ядровой трюк. Теорема Мерсера. Примеры ядер, методы построения ядер. Ядровой метод опорных векторов.
17. Наивный байесовский алгоритм, метод k ближайших соседей. Калибровка вероятностей.
18. Простые методы кодирования текстовых данных (Bag of Words, Tf-Idf, N-grams, хэширование), word2vec.
19. Методы поиска аномалий: правило трех сигм, интерквартильный размах, алгоритмы машинного обучения.
20. Методы отбора признаков: фильтрационные методы, оберточные методы, встроенные в модель методы.
21. Метод главных компонент (с выводом). Геометрическая интерпретация. Доля объясненной дисперсии.
22. Сингулярное разложение матрицы, его связь с методом главных компонент. Ядровой метод главных компонент.
23. Решающие деревья: определение и жадный алгоритм обучения. Функционал качества при выборе предиката. Общий вид критерия информативности (через функцию потерь). Критерии останова, обработка пропусков в данных.
24. Решающие деревья: критерий информативности и его значение. Общий вид критерия информативности (через функцию потерь), критерии информативности в задачах регрессии, классификации с предсказанием классов и классификации с предсказанием вероятностей классов.
25. Композиции алгоритмов. Разложение ошибки на смещение и разброс (с выводом).
26. Бэггинг и случайные леса. Смещение и разброс для бэггинга (в сравнении с одним базовым алгоритмом, с выводом).
27. Бустинг. Обучение базовых алгоритмов для регрессии с MSE (с выводом). Смещение и разброс у бустинга (в сравнении с одним базовым алгоритмом). Стекинг и блендинг.
28. Градиентный бустинг. Обучение базовых алгоритмов для произвольной дифференцируемой функции потерь. Сокращение шага.
29. Современные имплементации градиентного бустинга – XGBoost, CatBoost, LightGBM, их особенности.
30. Задача прогнозирования временного ряда, примеры задач. Адаптивные модели краткосрочного прогнозирования: экспоненциальное скользящее среднее (вывод рекуррентного соотношения), модель Хольта, модель Уинтерса, модель Тейла-Вейджа.
31. Компоненты временного ряда. Стационарность, методы избавления от нестационарности. Простейшие эконометрические модели временных рядов (AR, MA, ARMA, ARIMA).
32. Кросс-валидация на временных рядах. Извлечение признаков в задаче прогнозирования временных рядов с помощью методов машинного обучения. Адаптивная селекция и адаптивная композиция.
33. Задача кластеризации. Методы: K-Means, графовые методы, иерархическая кластеризация. Внешние и внутренние метрики качества кластеризации.
34. Задача визуализации. Multidimensional scaling, t-SNE.

# 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. Чем гиперпараметры отличаются от параметров? Что является параметрами и гиперпараметрами в линейных моделях и в решающих деревьях?
9. Что такое регуляризация? Почему L1-регуляризация отбирает признаки?
10. Запишите формулу для линейной модели классификации. Что такое отступ?
11. Что такое точность и полнота?
12. Что такое AUC-ROC? Как построить ROC-кривую?
13. Запишите функционал логистической регрессии. Как он связан с методом максимума правдоподобия?
14. Запишите задачу метода опорных векторов для линейно неразделимого случая. Как функционал этой задачи связан с отступом классификатора?
15. В чём заключаются one-vs-all и all-vs-all подходы в многоклассовой классификации?
16. В чём заключается преобразование категориальных признаков в вещественные с помощью счётчиков?
17. Что такое N граммы и как они используются в задачах с текстовыми данными?
18. Что такое сингулярное разложение матрицы?
19. Запишите постановку задачи в методе главных компонент. Как решение связано с ковариационной матрицей?
20. Зачем нужно отбирать признаки и как это можно делать?
21. Что такое ядро?
22. Запишите определения полиномиального и гауссовского ядра.
23. Опишите жадный алгоритм обучения решающего дерева.
24. Почему с помощью решающего дерева можно достичь нулевой ошибки на обучающей выборке без повторяющихся объектов?
25. Как в общем случае выглядит критерий информативности? В чем его значение?
26. Что такое бэггинг?
27. Что такое случайный лес? Чем он отличается от бэггинга над решающими деревьями?
28. Как в градиентном бустинге обучаются базовые алгоритмы? Что такое сокращение шага?
29. Какие деревья (глубокие или нет) нужно использовать в случайном лесе и в бустинге? Почему?
30. Зачем нужен критерий Дики-Фуллера?
31. Как работает метод K-Means?
32. Как работает метод визуализации t-SNE?
33. Как выглядит формула для Tf-Idf?
34. Что такое isolation forest
35. Как выглядит формула модели экспоненциального скользящего среднего?
36. Как происходит кросс-валидация на временных рядах?