

1. Модель данных Python; Типы данных
2. Базовые операции над векторами в Numpy (Задание).
3. Анализ данных с помощью Pandas. Основные операции. Базовый пайплайн анализа данных.
4. Аналитическое решение. Решение методом градиентного спуска. Регуляризация
5. Логистическая регрессия - алгоритм, loss function, регуляризация.
6. KNN - алгоритм и его модификации.
7. Метрики качества классификации
8. Решающие деревья - критерий информативности, алгоритм построения дерева, деревья для задачи регрессии, критерий останова. Сравнение с линейными моделями.
9. Bias-variance decomposition, bootstrapping, bagging. Random Forest.
10. Gradient Boosting - алгоритм + продемонстрировать на примере задачи регрессии.
11. Блендинг, стекинг - продемонстрировать на примере.
12. Нормализация текстов - лемматизация, стемизация, разбиение текста на слова
13. Векторное представление текстов - bag of word, tf-idf вектор, поиск похожих текстов