

Лекция 1

Введение в машинное обучение. Постановка задачи, основные понятия. Критерии качества. Данные. Модели и алгоритмы обучения.

Лекция 2-3

Линейная регрессия, метод наименьших квадратов. Немного о градиентном спуске (уже не "сухо" о методе, а о внедрении в модель). Регуляризация. Функции потерь. Логистическая регрессия. Многоклассовая классификация. Масштабируемость линейных моделей.

Лекция 4

Метод k -ближайших соседей. Функции расстояния (Манхэттенская, Минковского, косинусное расстояние, расстояние Жаккара, дивергенция Кульбака-Лейблера). Поиск ближайших соседей. Графы близости.

Лекция 5

Деревья решений. Информативность в задачах регрессии и классификации. Особенности данных: категориальные признаки, пропуски. Регуляризация деревьев решений.

Лекция 6-7

Ансамблевые методы. Смещение, разброс. Случайный лес. Бустинг. Стекинг. Градиентный бустинг: основы, применение к задаче регрессии, обобщение на функции потерь. Темп обучения, важность признаков.

Лекция 8-9

Метрики классификации и регрессии. Выбор метрик, матрица ошибок, примеры на моделях, сравнение результатов. Точность и полнота, $f1$ -мера, $recall@k$, $precision@k$, AUC. MAE, MAPE. Оптимизация метрик. Кросс-валидация. Подбор гиперпараметров.

Лекция 10

Вероятностные модели. Принцип максимальной энтропии. Обобщенные линейные модели. Метод наивного Байеса.

Лекция 11

Введение в глубокое обучение. Нейронные сети: история развития, постановка задачи, основные определения. Прямое и обратное распространение ошибки. Градиент сложной функции. Базовые архитектуры. Функции активации. Немного о тонкостях обучения.