

Коллоквиум по ML

Tinkoff Generation

16-20 декабря 2019

Теоретические вопросы

Вам достаётся два случайных из них, вам нужно подготовить по ним ответ, желательно с какими-то записями. Оценивается в 2 балла (по баллу за каждый билет).

Тема №1 Задачи машинного обучения

Обучение с учителем. Обучение без учителя. Регрессия. Классификация. Кластеризация

Тема №2 Приближение функции многочленами

Задача, модель, функция потерь. Проблема переобучения на примере интерполяции. Зачем нужно разбиение выборки на тренировочную и тестовую часть. Кросс-валидация.

Тема №3 Линейная регрессия

Задача и модель линейной регрессии. Регуляризация.

Тема №4 Логистическая регрессия

Задача и модель бинарной логистической регрессии.

Тема №5 Метрики классификации

Метрики классификации: accuracy, precision, recall, F1, ROC-AUC. Небинарная классификация.

Тема №6 Кластеризация

Задача кластеризации. K-means. DBSCAN. Агломеративная кластеризация.

Тема №7 Уменьшение размерности

Задача уменьшения размерности. SVD. PCA. T-SNE.

Тема №8 Решающие деревья

Алгоритм построения решающего дерева. Кросс-энтропия.

Тема №9 Ансамблирование

Ансамбль алгоритмов. Бэггинг. Случайный лес. Градиентный бустинг.

Тема №10 Обработка текстов

Токенизация, Лемматизация, CountVectorizer, OneHotVectorizer, Стемминг.

Тема №11 Релевантность документа

Задача проверки на релевантность документа на запрос. TF-IDF.

Тема №12 Рекомендательные системы

Постановка задачи. Бейзлайны. Фильтрация на основе содержания (content-based approach).
Коллаборативная фильтрация.

Тема №13 Теория вероятности

Вероятность. Условная вероятность. Формула Байеса. Формула полной вероятности.
Распределения вероятностей. ММП. Наивный байесовский классификатор.