

Прикладные задачи анализа данных, Майнер ИАД

Вопросы к экзамену

1. Постановка задачи в рекомендательных системах. Приведите примеры рекомендательных систем. Приведите примеры рейтингов и их вычисления для разных задач.
2. Memory-based подход в рекомендательных системах. Вычисление сходства пользователей и айтемов. User-based рекомендации. Плюсы и минусы подхода.
3. Модели со скрытыми переменными. Постановка задачи и функционал для обучения. Связь с матричными разложениями. Методы обучения: SGD и ALS.
4. Implicit ALS. Сравнение с обычной моделью со скрытыми переменными.
5. Факторизационные машины: модель, функционал, методы обучения, связь с моделью со скрытыми переменными.
6. Точечные оценки. Несмешённость и состоятельность.
7. Методы получения точечных оценок: plug-in оценки, метод моментов, метод максимального правдоподобия.
8. Доверительные интервалы: определение, примеры вычисления.
9. Проверка статистических гипотез. Определение, p-value, общий метод проверки.
10. Параметрические тесты и t-тест.
11. Непараметрические тесты, критерий Манна-Уитни.
12. Задача прогнозирования временных рядов. Основные компоненты временного ряда. Приведите примеры временных рядов.
13. Аддитивная и мультипликативная модели временных рядов. Подходы к прогнозированию временных рядов.
14. Задача поиска аномалий. Статистические методы и непараметрические оценки плотности. Isolation forest.
15. Задача извлечения именованных сущностей. Основные инструменты и подходы для извлечения именованных сущностей из текста.
16. Обучение с подкреплением: постановка задачи, понятие агента, награды, среды и их связь. Value и Q функции и их физический смысл. Как найти V и Q функции в самом простом случае? Что такое метод кросс-энтропии и как он работает?