

Экзаменационные вопросы и теормин по курсу “Методы Машинного Обучения”

группа 521, 2015 год

лекции читали Майсурадзе А. И. (Арчил Иверьевич) и Сенько О. В. (Олег Валентинович)

Экзаменационные вопросы

Основы анализа данных (часть Майсурадзе)

1. Место и роль ИАД в современной структуре человеческой деятельности. Три уровня технологий анализа данных, их назначение. Понятие о моделировании реального мира в науке. Физическая модель. Модель “решателя”. Информационная модель. Эвристическая модель. Основной приём ИАД для связи предметных областей и фундаментальной математики (обучение и эксплуатация эвристической информационной модели).
2. Основные модели данных (dataframe, multidimensional, similarity tensor, transactional). Гомогенные и гетерогенные модели. Фундаментальные задачи ИАД и основные инструменты статистики.
3. Распределение фундаментальных задач ИАД и основных инструментов статистики по моделям данных: в разрезе исходных данных, в разрезе результатов.

Основные модели данных (часть Майсурадзе)

4. Модель данных «признаковое описание объектов». Понятие о шкалах значений атрибутов. Представление реляционными технологиями. Схемы «звезда» и «снежинка».
5. Многомерная модель данных. Группирование объектов как переход к многомерной модели данных. Аналитические пространства. Измерения и категории. Показатели. Детализация. Функции агрегирования, типы показателей по агрегированию.
6. Транзакционная модель данных. Связанные с ней задачи.

Классификация как задача (часть Майсурадзе)

7. Общая задача классификации. Понятие об обучении и использовании. Объект, модель, алгоритм-классификатор. Универсальные ограничения. Локальные ограничения. Оптимизационный подход.
8. Функционалы качества на размеченной выборке. Частотные функционалы качества. Случай бинарной классификации. Стоимостные функционалы качества. Несоответствие частотных и стоимостных функционалов качества человеческому поведению.

Статистический подход к распознаванию (часть Сенько)

9. Понятие байесовского классификатора как оптимального алгоритма распознавания.
10. Классификаторы, основанные на использовании формулы Байеса. Линейный дискриминант Фишера.
11. Классификаторы, основанные на использовании формулы Байеса. Логистическая регрессия.
12. Метод k-ближайших соседей.

Практические навыки работы с данными (часть Майсурадзе)

13. Форматы представления информации. Текстовые файлы, их атрибуты, проблема определения атрибутов. Текстовые форматы представления таблиц: separated values, delimited text. Экранирование символов. Форматы представления транзакционных данных.
14. Диаграммы для многомерной модели данных. Системы отчётности.

Регрессионный анализ (часть Сенько)

15. Простая регрессия. Множественная регрессия. Поиск коэффициентов по МНК. Недостатки МНК.
16. Трёхкомпонентное разложение ошибки регрессионных моделей.
17. Регуляризация по Тихонову. Гребневая регрессия, лассо, эластичные сети.

Метрические тензоры (часть Общая)

18. Тут везде нужна только постановка задачи кластеризации:
Модель данных «метрические тензоры», гомогенные и гетерогенные многомерные матрицы сходства. Группирование объектов как кластеризация по метрическим описаниям. Гомогенная кластеризация, бикластеризация, мультикластеризация. Основные типы результатов кластеризации (плоская, последовательная плоская, иерархическая, нечёткая, стохастическая, ранговая).
19. Плоская кластеризация. Задача и метод k-means.

20. Последовательная плоская кластеризация. Метод ФОРЕЛЬ.
21. Иерархическая кластеризация. Дивизивная. Агломеративная, функционалы связи (linkage).

Модели и методы ИАД (часть Сенько)

22. ROC анализ.
23. Линейная модель. Линейная машина как метод обучения линейной модели.
24. Метод опорных векторов.
25. Нейронные сети. Модель перцептрона Розенблатта. Метод его обучения. Теорема Новикова. Переход от сдвига к фиктивному признаку. Многослойные перцептроны. Метод обратного распространения ошибки. Функции активации, удобные для распространения ошибки. Возможность разделения любых множеств.
26. Решающие деревья.
27. Комбинаторно логические методы. Тестовый алгоритм.
28. Основы АВО. Алгоритмы типа КОРА.
29. Алгоритмы, основанные на голосовании по наборам закономерностей.
30. Bagging. Boosting. Решающие леса (random forest).
31. Анализ выживаемости как задача.

Теормин

1. Понятие о фундаментальных задачах ИАД (основные задачи машинного обучения). Принципы их группирования.
2. Основные модели данных в ИАД. Признаковое описание объекта.
3. Основные модели данных в ИАД. Многомерная модель.
4. Основные модели данных в ИАД. Многомерная модель. Понятие и видах агрегируемости показателей.
5. Основные модели данных в ИАД. Тензоры сходства, графы взаимодействия.
6. Основные модели данных в ИАД. Транзакционные данные. Формальные контексты.
7. Фундаментальные задачи ИАД. Задача классификации. Задача восстановления регрессии.
8. Фундаментальные задачи ИАД. Задача кластерного анализа.
9. Фундаментальные задачи ИАД. Задача сокращения размерности.
10. Обобщающая способность.
11. Принцип минимизации эмпирического риска.
12. Методы оценивания обобщающей способности. Скользящий контроль.
13. Регрессионный анализ. Метод наименьших квадратов (МНК). Простая одномерная регрессия.
14. Множественная линейная регрессия. Недостатки МНК.
15. Использование методов регуляризации.
16. Методы распознавания. Что такое байесовский классификатор? Байесовские методы в распознавании.
17. Метод k-ближайших соседей.
18. Комбинаторно-логические методы. Тупиковый тест. Представительный набор.
19. Алгоритм вычисления оценок.
20. Модель нейрона. Перцептрон Розенблатта.
21. Многослойный перцептрон. Методы обучения.
22. Метод опорных векторов. Основные понятия. Возможность нелинейного разделения.
23. Методы решающих деревьев и решающих лесов.
24. Методы построения коллективных решений в распознавании.
25. Основы алгебраической коррекции.
26. Задача анализа выживаемости (краткое определение)