

Вопросы к экзамену по курсу

«Алгебраические методы обработки данных», осень 2015

1. Булевы функции, их запись, изображения на булевом кубе.
2. Дизъюнктивные нормальные формы (ДНФ): сокращённые, тупиковые, кратчайшие.
3. Алгоритмы построения ДНФ: метод Нельсона, метод Блейка, критерий поглощения.
4. Алгоритмы вычисления оценок (АВО), обобщения АВО, эффективные формулы для оценок.
5. Алгебра над алгоритмами, линейное и алгебраическое замыкание.
6. База в линейном замыкании АВО.
7. Постановка задачи распознавания по прецедентам. Сущность дискретного (логического) подхода к задачам распознавания.
8. Общие принципы построения дискретных (логических) процедур распознавания. Понятие корректного элементарного классификатора для случая целочисленных данных.
9. Модели тестовых алгоритмов распознавания и используемые в них семейства элементарных классификаторов.
10. Модели алгоритмов голосования по представительным наборам и используемые в них семейства элементарных классификаторов.
11. Модели алгоритмов голосования по покрытиям классов и используемые в них семейства элементарных классификаторов.
12. Построение элементарных классификаторов для тестовых алгоритмов распознавания и алгоритмов голосования по представительным наборам на основе поиска покрытий булевых матриц.
13. Построение элементарных классификаторов в алгоритмах голосования по представительным наборам на основе преобразования нормальных форм логических функций (на примере бинарных признаков).
14. Задача дуализации (формулировка с использованием понятий теории нормальных форм булевых функций). Основные подходы к оценке эффективности алгоритмов дуализации в «худшем случае».
15. Задача дуализации (формулировка с использованием понятия неприводимого покрытия булевой матрицы). Асимптотически оптимальные алгоритмы дуализации.
16. Алгебро-логический подход к построению корректных процедур распознавания на базе произвольных (не обязательно корректных) элементарных классификаторов. Понятие корректного набора элементарных классификаторов.
17. Общая схема работы логического корректора. Подходы к снижению вычислительной сложности на этапе обучения логического корректора. Практические модели логических корректоров.
18. Методы повышения эффективности дискретных (логических) процедур распознавания. Оценка информативности признаков, значений признаков, выделение шумящих признаков и обучающих объектов, не являющихся типичными для своего класса.
19. Основные модели данных (dataframe, multidimensional, similarity tensor, transactional), краткое определение. Гомогенные и гетерогенные модели. Распределение фундаментальных задач ИАД и основных инструментов статистики по моделям данных: в разрезе исходных данных, в разрезе результатов.

20. Модель данных «признаковое описание объектов». Понятие о шкалах значений атрибутов. Представление реляционными технологиями. Схемы «звезда» и «снежинка». Диаграммы для наборов точек из R^n .
21. Многомерная модель данных. Группирование объектов как переход к многомерной модели данных. Аналитические пространства. Измерения и категории. Показатели. Детализация. Функции агрегирования, типы показателей по агрегированию. Соответствующие диаграммы. Системы отчётности.
22. 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, иерархическая, нечёткая, стохастическая, ранговая).
23. Задачи точной реализации метрических тензоров. Корректность задачи (разрешимость, однозначность). Алгоритмическая сложность (на примере метрик Минковского). Метрическое многомерное шкалирование, его связь с методом главных компонент.
24. Задачи аппроксимации метрических тензоров. Неметрическое многомерное шкалирование. Функционалы стресса. Монотонные (изотонические) отображения, сохранение ранга метрического тензора. Достаточная размерность представления для неразрешимых задач.
25. Задача кластеризации как задача аппроксимации метрического тензора. Метрики на метриках, аппроксимация метрик метриками. Метрические деревья. Ультраметрические деревья. Филогенетические деревья, интерпретация длин ребёр и нетерминальных вершин в ультраметрических деревьях, гипотеза молекулярных часов. Гарантированное получение классов эквивалентности. Общая схема вычисления ближайшей ультраметрики.
26. Трёхкомпонентное разложение ошибки. Bias-Variance дилемма.
27. Разложение ошибки для выпуклых комбинаций предикторов. Несократимые комбинации. Разложение ошибки для компоненты сдвига и вариационной компоненты обобщённой ошибки.
28. Методы верификации закономерностей, основанные на перестановочных тестах. Метод оптимальных достоверных разбиений.
29. Метод континуального голосования в модели АВО.
30. Метод статистически взвешенных синдромов.