Кафедра ИИТ + МС

Вопросы к государственному экзамену
Магистерская программа «Интеллектуальный анализ больших данных»

1. Линейная регрессионная модель. Метод наименьших квадратов. Основные свойства оценок метода наименьших квадратов.
2. Проверка линейных гипотез в рамках классической модели регрессии.
3. Дискриминантный анализ: постановка задачи и ее решение в случае известных параметров.
4. Кластерный анализ: постановка задачи и основные понятия.
5. Спектральный анализ временных рядов. Периодограмма выборочная. Свойства периодограммы. Сглаженная периодограмма.
6. Линейная ARMA(p,q) модель временных рядов. Стационарность, обратимость. Автокорреляционная и частная автокорреляционная функции. Оценивание параметров. Прогнозирование.
7. Общие принципы выбора страховых тарифов и тарифных ставок.
8. Асимптотическая формула для страховых тарифов в статической модели страхования: сравнение классической и факторизационной модели.
9. Гарантированные оценки страховых тарифов для факторизационной модели страхования при пуассоновском объеме портфеля.
10. Преобразование Фурье и его свойства. Линейные фильтры. Теорема о свертке.
11. Кратномасштабный анализ. Алгоритмы разложения и реконструкции.
12. Пороговая обработка коэффициентов вейвлет-разложения. Методы выбора порога.
13. Модель вычислений Map Reduce. Пример реализации (GoogleMapReduce либо HadoopMapReduce). Стадии вычислений, основные компоненты, понятие локальности данных.
14. Модель обработки данных в Apache Spark. Понятие resilient distributed dataset (RDD), примеры преобразований и действий над RDD. Граф вычислений.
15. Распределенная файловая система. Пример распределенной файловой системы (Google File System либо HDFS). Компоненты, основные операции, сценарии чтения и записи данных, механизмы отказоустойчивости.
16. Способы объектно-реляционного отображения для классов и атрибутов, бинарных и N-арных ассоциаций, классов ассоциаций, иерархий наследования.Примеры применения этих способов. Моделирование схемы реляционной базы данных с помощью диаграммы классов.
17. Образцы (паттерны) проектирования, их классификация и способ описания. Примеры образцов: структурного, поведенческого и порождающего.
18. Модели прогнозирования на основе деревьев решений. Алгоритмы CHAID, CART, C4.5: критерии поиска разбиений, параметры ограничения роста и обрубания дерева.
19. Нейронные сети прямого распространения. Архитектуры MLP и RBF: структура сетей, виды функций активации, алгоритмы обучения, борьба с переобучением и с проблемой локальных минимумов.
20. Поиск ассоциативных правил. Алгоритмы Apriori и FP-tree: построение частых эпизодов с ограничением по поддержке и формирование правил с ограничением по достоверности.
21. Векторная модель информационного поиска. Показатели tf и idf. Вычисление близости между запросом и документом. Применение метода "Наивный Байес "для автоматической классификации текстов.
22. Алгоритм ссылочного ранжирования Page Rank: способ вычисления, роль в информационном поиске.
23. Основные структуры данных в языке R: векторы, матрицы, массивы, факторы, списки, дата фреймы. Подготовка и манипулирование данными с помощь дата фреймов: отбор/удаление переменных, наблюдений, объединения наборов.
24. Применение функции lm для построения, интерпретации и диагностики одно- и многофакторных моделей дисперсионного анализа.
25. Применение функции glm для построения и анализа обобщенных линейных моделей. Построение, интерпретация и диагностика логистической, гамма и пуассоновской регрессионных моделей.
26. Преобразование изображения на основе эквализации гистограммы яркости.
27. Свёртка функций. Одномерная и двумерная свёртка и её свойства. Дискретная свертка изображений. Обработка края изображения при свёртке.
28. Последовательная и параллельная сложность алгоритмов, информационный граф и ресурс параллелизма алгоритмов.
29. Архитектурные особенности графических процессоров, направленные на массивно-параллельные вычисления.

Списокрекомендованнойлитературы

1. АндерсонТ.Введениевмногомерныйстатистическийанализ.–М.:Наука,1963.
2. ШеффеГ.Дисперсионныйанализ.–М.:Наука, 1980.
3. HardleW.,SimarL. AppliedMultivariateStatisticalAnalysis.–Springer-Verlag,2003.
4. JohnsonR.A.andWinchernD.W.AppliedMultivariateStatisticalAnalysis.–PearsonPrenticeHall, 2007.
5. Факторный, дискриминантный и кластерный анализы. – М.: Финансы истатистика,1989.
6. СеберДж.Линейныйрегрессионныйанализ.М.:Мир,1980.
7. ДженкинсГ.ВаттсД.Спектральныйанализиегоприложения.М.Мир,1972.
8. Бокс Дж., Дженкинс Г. Анализ временных рядов прогноз и управление. М. Мир,1974.
9. Б.В.Гнеденко, Э.А.Даниелян, Б.Н.Димитров, Г.П.Климов,В.Ф.Матвеев,Приоритетные системы обслуживания. Издательство Московского университета,1973.
10. В.Ф.Матвеев,В.Г.УшаковСистемымассовогообслуживания.ИздательствоМосковскогоуниверситета, 1984.
11. Захарова Т.В., Шестаков О.В. Вейвлет-анализ и его приложения. 2-е изд., перераб.идоп. Учебноепособие. М.: ИНФРА-М.2012.
12. ЧакЛэм.Hadoopвдействии.–М:ДМКПресс.–2012,424с.
13. Х.Карау,Э.Конвински,П.Венделл,М.Захария.ИзучаемSpark.Молниеносныйанализданных.– М: ДМК Пресс.– 2015, 304с.
14. РамбоДж.,БлахаM.UML2.0.Объектно-ориентированноемоделированиеиразработка.– СПб.:Питер.–2007. Главы 3 и 19.
15. ГаммаЭ.идр.Приемыобъектно-ориентированногопроектирования.Паттерныпроектирования.:Пер. сангл.–СПб.:Питер, 2016.

16. ГонсалесР.,ВудсР.Цифроваяобработкаизображений.:Пер.сангл.– М.:Техносфера,2006.– 1070 с.
17. ШапироЛ.,СтокманДж.Компьютерноезрение:Пер.сангл.– М.:БИНОМ.Лабораториязнаний, 2006.– 752 с.
18. Hastie,TibshiraniandFriedman.TheElementsofStatisticalLearning. –Springer-Verlag,2009.– 763pages.http://statweb.stanford.edu/~tibs/ElemStatLearn/
19. СаймонХайкин.Нейронныесети.Полныйкурс.2-еизд.,испр.:Пер.сангл.–М.:ООО«И. Д.Вильямс»,2006.– 1104 с.
20. LoraD.Delwiche,SusanJ.Slaughter.TheLittleSAS®Book:APrimer,FifthEdition. -SASInstitute,2012. -331pages.
21. SAS/STATUser'sGuide. http://support.sas.com/documentation/onlinedoc/stat/index.html
22. МаннингК.,РагхаванП.,ШютцеХ.Введениевинформационныйпоиск.— Вильямс,2011.
23. ЛукашевичН.В.Тезаурусывзадачахинформационногопоиска.–М.:Изд-воМосковскогоуниверситета, 2011
24. ИнгерсоллГ.,МортонТ.Обработканеструктурированныхтекстов.Поиск,организация манипулирование.ДМК-Пресс,2015.