

Материалы для подготовки к тестированию компетенций

DataScientists

МАТЕМАТИКА

Математический анализ

Темы:

1. Сумма и сходимость рядов
2. Интегрирование элементарных функций одной переменной.
3. Дифференцирование элементарных функций.
4. Задачи оптимизации на отрезке (поиск оптимума, условия экстремума первого и второго порядка)

Литература, онлайн курсы:

1. Зорич В. А. Математический анализ, ч. 1.
2. Демидович БП. Сборник задач и упражнений по математическому анализу. АСТ; 2006.

Дискретная математика

Темы:

1. Булева алгебра. Конъюнкция, дизъюнкция, импликация, отрицание. Их математические обозначения и свойства.
2. Теория множеств: пересечение, объединение, дополнение. Их математические обозначения и свойства.
3. Теория графов. Обход графа в глубину и в ширину. Поиск кратчайшего пути.

Литература, онлайн курсы:

- Кормен Т., Лейзер Ч.: Алгоритмы. Построение и анализ. (I.5, V.23)
- Кузнецов О.П., Адельсон-Вельский Г.М.: Дискретная математика для инженера.
- ШАД. Курс лекций «Дискретный анализ и теория вероятностей»
<https://yandexdataschool.ru/edu-process/courses/probability>

Алгоритмы

Темы

- Алгоритмы сортировки и поиска
- Сложность алгоритмов

Литература, онлайн курсы:

- Кормен Т., Лейзер Ч.: Алгоритмы. Построение и анализ. (Введение, II, III)
- ШАД. Курс лекций «Алгоритмы и структуры данных» <https://yandexdataschool.ru/edu-process/courses/algorithms>

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Основы статистического анализа

Темы:

1. Комбинаторика и основы теории вероятностей
2. Число сочетаний, перестановок.
3. Определение вероятности. Произведение и сумма вероятностей, условная вероятность.
4. Дискретные случайные величины: бернулли, биномиальное, пуассоновское распределения.
5. Непрерывные случайные величины: равномерное, экспоненциальное, нормальное, логнормальное распределение.
6. Математическое ожидание, дисперсия, ковариация, корреляция, квантили, коэффициент эксцесса, коэффициент асимметрии.

Литература, онлайн курсы:

- Кормен Т., Лейзер Ч.: Алгоритмы. Построение и анализ. I.6 Комбинаторика и вероятность
- <https://yandexdataschool.ru/edu-process/courses/probability> 1, 10 лекции курса ШАД "Дискретный анализ и теория вероятностей"

Эконометрика и статистика

Темы:

1. Линейная регрессия. Теорема Гаусса-Маркова. Свойства оценок. Тестирование гипотез (t-test), p-value.
2. Метод максимального правдоподобия. Регуляризация в линейной регрессии

Литература, онлайн курсы:

- Гнеденко Б. В.: Курс теории вероятностей.

- <https://yandexdataschool.ru/edu-process/courses/machine-learning>

Машинное обучение.

Временные ряды

Темы:

1. Определение трендов, сезонности, периодичности
2. Предсказание временного ряда. Сглаживание временного ряда.

Литература, онлайн курсы:

- Бокс Г., Дженкинс Г.: Анализ временных рядов: Прогноз и управление.
- Лоскутов А.Ю.: Анализ временных рядов. М.: Физич. фак-т МГУ им. МВ Ломоносова. 2009.

МАШИННОЕ ОБУЧЕНИЕ

Машинное обучение I

Темы:

1. Подготовка данных
2. Оценка и мониторинг моделей
3. Обучение без учителя
4. Задача классификации
5. Деревья и ансамбли из них
6. Линейные модели / SVM
7. Методы понижения размерности и отбор признаков

Литература, онлайн курсы:

- Воронцов К. В. Машинное обучение (курс лекций) //Электронная версия: http://bit.ly/ml_course – 2009, его сайт <http://www.machinelearning.ru>
- <https://ru.coursera.org/specializations/machine-learning-data-analysis> Специализация МФТИ на Coursera
- Документация scikit-learn: scikit-learn.org/stable/documentation.html

Машинное обучение II

Темы:

1. Ансамбли. Стеккинг. Бегинг. Бустинг.
2. Нейронные сети и глубинное обучение. Обучение, алгоритмы оптимизации. Сверточные нейронные сети. Рекуррентные нейронные сети.
3. Рекомендательные системы. Различные подходы к
4. Автоматическая обработка текстов. Лемматизация, стемминг. Определение части речи. Строковые меры близости. Классификация текстов.
5. Ассоциативные правила. Алгоритм Apriori.
6. Компьютерное зрение. Сегментация и классификация изображений.

Литература, онлайн курсы:

- Воронцов К. В. Машинное обучение (курс лекций) //Электронная версия: http://bit.ly/ml_course – 2009.
- <https://ru.coursera.org/specializations/machine-learning-data-analysis> Специализация МФТИ на Coursera
- Форсайт Д, Понс Ж. Компьютерное зрение. Современный подход. М.: ИД Вильямс; 2004.
- Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В.: Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие. М.: МИЭМ. 2011;272:3.
- Николенко С.И., Кадури А.А., Архангельская Е.О.: Глубокое обучение. Издательский дом Питер, 2017

ПРОГРАММИРОВАНИЕ И ТЕХНОЛОГИИ РАБОТЫ С ДАННЫМИ

Темы:

1. Python / R / SAS
2. Hadoop, MapReduce
3. SQL

Краткое ознакомление:

- http://www.sql.ru/docs/sql/u_sql/
- <https://habr.com/post/240405/>
- <http://xgu.ru/wiki/Hadoop>