

Математика для анализа данных

01 Кому подойдёт курс

Начинающим аналитикам

Чтобы закрыть пробелы в статистике, понимать контекст применения статистических тестов и освоить методы визуализации больших данных.

Начинающим специалистам по Data Science

Чтобы разобраться, что «под капотом» у знакомых инструментов, и с лёгкостью осваивать новые.

Тем, кто готовится к собеседованиям в IT-компаниях

Чтобы подтянуть нужные знания и быть готовым к любым вопросам, даже если со школы не вспоминали математику.

Тем, кто недавно окончил курсы по анализу данных

Чтобы укрепить навыки и претендовать на вакансии, где ценят хорошее знание математики.

02 Чему научитесь за 6 месяцев

- A/B-тесты, стат. тесты, доверительный интервал, p-value
- Линейная регрессия и сингулярное разложение
- Градиентный спуск и другие алгоритмы обучения нейросетей
- Косинусное расстояние между текстами

03 Как проходит курс

- Теория и практика на платформе Практикума
- Доступ из любой точки мира в удобное время
- Воркшопы
- Практические задания

Что вас ждёт

01

Объяснение сложного простым языком

Подробно разбираем каждое понятие с примерами и иллюстрациями

02

Помощь в освоении навыков для работы и собеседований

03

Много практических задач и бизнес-кейсов

Чтобы навык закрепился, каждое занятие завершаем практикой. Показываем, как решать математические задачи на Python

Математика для анализа данных

00

Вводная часть

- Связь математики и современных инструментов анализа данных
- Разбор кейсов и задач, в которых аналитикам нужна математика

01

Линейная алгебра

- Векторы
- Нормы
- Матрицы
- Обратная матрица и определитель

02

Функции и их свойства

- Линейная функция, Полином, Логарифм
- Модуль, Композиция

03

Математический анализ

- Производные и интегралы
- Функции нескольких переменных
- Градиентный спуск

04

Приложения линейной алгебры в анализе данных

- Линейная регрессия
- Сингулярное разложение, PCA, визуализация данных высокой размерности и другие применения SVD

05

Теория вероятностей и основы статистики

- Дискретные случайные величины
- Взаимодействие дискретных величин
- Непрерывные случайные величины

06

Статистические методы

- Статистическая оценка параметров
- Статистические эксперименты и проверка гипотезы
- Методы статистической проверки гипотез
- Метод главных компонент



Собеседование

- Симулятор математической секции собеседования

Время прохождения:
в среднем 1 час

Узнаете, для понимания каких инструментов анализа данных и Data Science нужен каждый раздел математики. Разберёте несколько математических кейсов, которые встречаются на собеседованиях и в работе аналитиков.

Содержание

1. Как применяют математику в анализе данных
2. Кейсы попроще
3. Кейсы посложнее

Линейная алгебра

01

5 недель,
в среднем 50 часов

Научитесь использовать матрицы, векторы, нормы, определители. Сможете читать обозначения и оперировать формулами. Узнаете, почему косинусное расстояние используется для сравнения текстов. Разберётесь в особенностях применения линейной алгебры в анализе данных.

Содержание

1 неделя,
в среднем 10 часов

Векторы

1. Операции над векторами
2. Векторное пространство
3. Основы тригонометрии

1 неделя,
в среднем 10 часов

Нормы

1. Скалярное произведение векторов
2. Нормы вектора
3. Связь L2 нормы и скалярного произведения
4. Расстояния между векторами

1,5 недели,
в среднем 15 часов

Матрицы

1. Линейная (не)зависимость
2. Базис
3. Арифметические операции над матрицами и их свойства
4. Умножение матрицы на вектор
5. Матричное перемножение

1,5 недели,
в среднем 15 часов

Обратная матрица и определитель

1. Обратная матрица
2. Вырожденная матрица
3. Определитель, его применение и смысл
4. Упрощение матричных выражений

2 недели,
в среднем 20 часов

Изучите виды функций, которые часто встречаются в аналитике данных и Data Science. Разберётесь с понятием функции и её графиком, рассмотрите линейную и полиномиальную функции. Выясните, как с их помощью аппроксимировать данные.

Содержание

1 неделя,
в среднем 10 часов

Функции, часть 1

1. Определение функции и графика функции
2. Линейная функция
3. Полиномиальная функция
4. Описание данных с помощью функций

1 неделя,
в среднем 10 часов

Функции, часть 2

1. Показательная функция
2. Логарифм
3. Обратная функция
4. Модуль
5. Композиция функций

Математический анализ

03

5 недель,
в среднем 50 часов

Освойте базу, которая лежит в основе работы многих инструментов. Например, узнаете про поиск оптимального решения с помощью градиентного спуска.

Содержание

2 недели,
в среднем 20 часов

Производная

1. Предел
2. Производная
3. Правила нахождения производных
4. Экстремумы функции
5. Свойства функций: монотонность, выпуклость

1,5 недели,
в среднем 15 часов

Интегралы

1. Первообразная и неопределённый интеграл
2. Определённый интеграл
3. Кусочно-заданные функции и их интегрирование
4. Интегрирование по частям и замена переменных
5. Несобственный интеграл

1.5 недели,
в среднем 15 часов

Функции нескольких переменных

1. Функция нескольких переменных
 2. Визуализация функции двух переменных
 3. Частная производная
 4. Градиент функции
 5. Экстремумы функции нескольких переменных
 6. Градиентный спуск
-

Приложения линейной алгебры в анализе данных

04

3 недели,
в среднем 30 часов

Поймёте, как работают методы линейной регрессии и сингулярного разложения. Узнаете, как связаны собственные числа с матричными разложениями PCA и SVD, и научитесь их вычислять. Научитесь сокращать размерность больших данных и визуализировать их. Узнаете, как найти решение линейной регрессии при помощи градиентного спуска, и лучше поймёте, как обучается нейронная сеть.

Содержание

1.5 недели,
в среднем 15 часов

Линейная регрессия

1. Определение модели линейной регрессии
 2. Нахождение параметров линейной регрессии как решение СЛУ
 3. Функция ошибки и её минимизация
 4. Аналитическое решение для параметров регрессии
 5. Коллинеарность в данных, регуляризация
 6. Градиентный спуск для нахождения параметров линейной регрессии
-

1.5 недели,
в среднем 15 часов

Сингулярное разложение, визуализация данных большой размерности

1. Собственные значения и векторы матрицы
 2. Геометрическое представление собственных векторов
 3. Сингулярные векторы, сингулярное разложение SVD
 4. Применения сингулярного разложения
-

Теория вероятностей и основы статистики

05

4 недели,
в среднем 40 часов

Научитесь работать с дискретной и непрерывной случайной величиной, вычислять корреляцию и ковариацию. Разберётесь в теореме Байеса. Изучите часто используемые распределения: биномиальное, равномерное, нормальное и другие. Освойте понятия выборки и генеральной совокупности.

Содержание

1 неделя,
в среднем 12 часов

Дискретные случайные величины

1. Исход, событие, вероятность
2. Дискретные случайные величины и их свойства
3. Дискретные распределения

1.5 недели,
в среднем 14 часов

Взаимодействие дискретных случайных величин

1. Совместное распределение и ковариация
2. Условная вероятность, теорема Байеса
3. Сэмплирование, базовая визуализация данных
4. Основные описательные статистики

1.5 недели,
в среднем 14 часов

Непрерывные случайные величины

1. Непрерывные случайные величины
2. Нормальное распределение и ЦПТ
3. Совместное распределение, связь величин, корреляция
4. Условная вероятность и теорема Байеса для непрерывных величин
5. Гистограммы и описательные статистики для непрерывных величин

Статистические методы

06

5 недель,
в среднем 50 часов

Узнаете, как принимаются решения на основе статистических данных. Научитесь выбирать способы оценивания данных, которые минимизируют ошибку. Узнаете, что такое доверительный интервал и бутстреп. Освойте математическую базу A/B-тестирования и метод главных компонент.

Содержание

1,5 недели,
в среднем 14 часов

Статистическая оценка параметров

1. Вероятность и правдоподобие
2. Оценка параметров распределения
3. Метод максимального правдоподобия
4. Поиск параметров линейной регрессии с вероятностной точки зрения

1,5 недели,
в среднем 14 часов

Статистические эксперименты и проверка гипотез

1. Вероятностное принятие решений
2. Доверительные интервалы
3. A/B-тестирование
4. Параметрические тесты
5. Размер выборки, его связь с ошибкой

1 неделя,
в среднем 12 часов

Методы статистической проверки гипотез

1. Непараметрические тесты
2. Бутстреп
3. Нелинейное преобразование данных
4. Множественная проверка гипотез

1 неделя,
в среднем 10 часов

Метод главных компонент

1. Матрица ковариации
2. Применение PCA
3. SVD для расчёта PCA

Симуляция математической секции собеседования



Дополнительный модуль

Пройдёте симулятор математической секции собеседования на позицию аналитика или специалиста по Data Science.

