

Математика для анализа данных



Время прохождения:
в среднем 1 час

0 Вводная часть

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.

- Как применяют математику в анализе данных
- Кейсы попроще
- Кейсы посложнее

5 недель,
в среднем 50 часов

① Линейная алгебра

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.

1 неделя,
в среднем 10 часов

Векторы

- Операции над векторами
- Векторное пространство
- Основы тригонометрии

1 неделя,
в среднем 10 часов

Нормы

- Скалярное произведение векторов
- Нормы вектора
- Связь L2 нормы и скалярного произведения
- Расстояния между векторами

1,5 недели,
в среднем 15 часов

Матрицы

- Линейная (не)зависимость
- Базис
- Арифметические операции над матрицами и их свойства
- Умножение матрицы на вектор
- Матричное перемножение

1,5 недели,
в среднем 15 часов

Обратная матрица и определитель

- Обратная матрица
- Вырожденная матрица
- Определитель, его применение и смысл
- Упрощение матричных выражений

2 недели,
в среднем 20 часов

② Функции и их свойства

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

1 неделя,
в среднем 10 часов

Функции, часть 1

- Определение функции и графика функции
- Линейная функция
- Полиномиальная функция
- Описание данных с помощью функций

1 неделя,
в среднем 10 часов

Функции, часть 2

- Показательная функция
- Логарифм
- Обратная функция
- Модуль
- Композиция функций

5 недель,
в среднем 50 часов

③ Математический анализ

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

2 недели,
в среднем 20 часов

Производная

- Предел
- Производная
- Правила нахождения производных
- Экстремумы функции
- Свойства функций: монотонность, выпуклость

Интегралы

1,5 неделя,
в среднем 15 часов

- Первообразная и неопределённый интеграл
- Определённый интеграл
- Кусочно-заданные функции и их интегрирование
- Интегрирование по частям и замена переменных
- Несобственный интеграл

1.5 недели,
в среднем 15 часов

Функции нескольких переменных

- Функция нескольких переменных
- Визуализация функции двух переменных
- Частная производная
- Градиент функции
- Экстремумы функции нескольких переменных
- Градиентный спуск

3 недели,
в среднем 30 часов

④ Приложения линейной алгебры в анализе данных

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 поймете, как обучается нейронная сеть.

1.5 недели,
в среднем 15 часов

Линейная регрессия

- Определение модели линейной регрессии
- Нахождение параметров линейной регрессии как решение СЛУ
- Функция ошибки и её минимизация
- Аналитическое решение для параметров регрессии
- Коллинеарность в данных, регуляризация
- Градиентный спуск для нахождения параметров линейной регрессии

1.5 недели,
в среднем 15 часов

Сингулярное разложение, визуализация данных большой размерности

- Собственные значения и векторы матрицы
- Геометрическое представление собственных векторов
- Сингулярные векторы, сингулярное разложение SVD
- Применения сингулярного разложения

4 недели,
в среднем 40 часов

5 Теория вероятностей и основы статистики

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. Освоите понятия выборки и генеральной совокупности.

1 неделя,
в среднем 12 часов

Дискретные случайные величины

- Исход, событие, вероятность
- Дискретные случайные величины и их свойства
- Дискретные распределения

1.5 недели,
в среднем 14 часов

Взаимодействие дискретных случайных величин

- Совместное распределение и ковариация
- Условная вероятность, теорема Байеса
- Сэмплирование, базовая визуализация данных
- Основные описательные статистики

1.5 недели,
в среднем 14 часов

Непрерывные случайные величины

- Непрерывные случайные величины
- Нормальное распределение и ЦПТ
- Совместное распределение, связь величин, корреляция
- Условная вероятность и теорема Байеса для непрерывных величин
- Гистограммы и описательные статистики для непрерывных величин

5 недель,
в среднем 50 часов

6 Статистические методы

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

1,5 неделя,
в среднем 14 часов

Статистическая оценка параметров

- Вероятность и правдоподобие
- Оценка параметров распределения
- Метод максимального правдоподобия
- Поиск параметров линейной регрессии с вероятностной точки зрения

1,5 неделя,
в среднем 14 часов

Статистические эксперименты и проверка гипотез

- Вероятностное принятие решений
- Доверительные интервалы
- А/В тестирование
- Параметрические тесты
- Размер выборки, его связь с ошибкой

1 неделя,
в среднем 12 часов

Методы статистической проверки гипотез

- Непараметрические тесты
- Бутстреп
- Нелинейное преобразование данных
- Множественная проверка гипотез

1 неделя,
в среднем 10 часов

Метод главных компонент

- Матрица ковариации
- Применение PCA
- SVD для расчёта PCA

Симуляция математической секции собеседования

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