

# Математика для Data Science

Best Practice по изучению математического аппарата, необходимого для успешной карьеры в Data Science

Длительность курса: 138 академических часов

## 1 Линейная алгебра

### 1 Математика в DataScience

вы поймете, где же именно в Data Science встречается математика и почему знание математики — это необходимое условие, чтобы стать успешным в этой области.

Домашние задания

- 1 Сложение и умножение матриц
  - 2 Посчитать собственные числа и вектора.
-

**2 Основные термины и определения математического анализа, линейной алгебры и теории вероятностей**

- 1) базовые термины матанализа:
  - предел;
  - непрерывность функции;
  - дифференциал;
- 2) базовые термины линейной алгебры:
  - вектор;
  - матрица, ее виды;
  - ранг;
  - определитель;
- 3) базовые термины теории вероятности:
  - определение вероятности;
  - мат. ожидание;
  - дисперсия;
- 4) установка Python, предоставление рекомендаций по интерпретатору;
- 5) вычисление базовых задач на Python с помощью стандартных пакетов.

Домашние задания

- 1 Сложение и умножение матриц.
- 

**3 Матрицы. Основные понятия и операции**

- ключевые определения;
- операция над матрицами;
- определитель;
- обратная матрица;
- вычисления собственных значений и собственных векторов;
- квадратичные формы.

Домашние задания

- 1 Посчитать собственные числа и вектора.
-

4	<b>Геометрическая интерпретация в линейной алгебре</b>	<ul style="list-style-type: none"><li>— геометрическая интерпретация матричных преобразований;</li><li>— правило Крамера;</li><li>— матрицы;</li><li>— матрица Маркова;</li><li>— форма.</li></ul> <hr/>
5	<b>Матричные разложения</b>	<p>Разложение SVD и прочие.</p> <p>Домашние задания</p> <p>1 Разложить матрицу в SVD.</p> <hr/>
6	<b>Матричные производные</b>	<ul style="list-style-type: none"><li>— матричные производные;</li><li>— дифференциальные уравнения в матрицах.</li></ul> <hr/>
7	<b>Применение линейной алгебры в Data Science.</b>	<p>классификация данных с SVM.</p> <p>Домашние задания</p> <p>1 Повторить вычисления с занятия в Python на других данных.</p> <hr/>
8	<b>Применение линейной алгебры в Machine Learning</b>	<p>распознавание изображений.</p> <hr/>

## 9 MidTerm

Проверочная работа. В ходе занятия будет предоставляться от 2-х до 5-ти минут на решение практической задачи. По истечению отведенного времени преподаватель решает задание на экране.

- 1 **Теория множеств**
- вероятностные пространства;
  - дискретное пространство элементарных исходов;
  - вероятность на числовой прямой и плоскости;
  - правило сложения и умножения.
- 

- 2 **Метрические пространства**
- понятие метрического пространства;
  - определение нормированного пространства, понятие нормы, отличие от метрики, примеры нормированных пространств;
  - норма в оптимизации.

### Домашние задания

- 1 Свойство метрики. Вычислить простые операции над множеством.
- 

- 3 **Теория пределов**
- определение Коши;
  - определение Пиано;
  - вычисление пределов функций;
  - асимптотические функции;
  - эквивалентные функции;
  - оценка сложности функции.
-

## 4 Дифференцирование

- дифференцируемость функции в точке;
- частные производные и дифференциалы высших порядков;
- градиент;
- матрица Гессе.

Домашние задания

- 1 Вычисление производных и пределов значений функции в бесконечности.
- 

## 5 Оптимизация

- экстремумы функций многих переменных;
  - определения точек локального и глобального минимума;
  - необходимое и достаточное условие экстремума для выпуклых функций;
  - понятие стационарных точек и отличие в их определении от точек экстремума.
- 

## 6 Минимизация и Максимизация в Регрессиях

- МНК;
- ММ.

Домашние задания

- 1 Максимизация функции с ограничениями. Минимизация квадрата ошибки.
- 

## 7 Интегрирование

- неопределенный интеграл;
  - определенный интеграл;
  - приложения определенного интеграла и приближенные методы его вычисления;
  - несобственные интегралы;
  - двойные интегралы;
  - приближенные методы интегрирования.
-

## 8 Поиск экстремумов в ML

- покоординатный спуск;
- градиентный спуск;
- адаптивные варианты градиентного спуска.

Домашние задания

- 1 Поиск экстремума с Python.

---

## 9 MidTerm

Проверочная работа. В ходе занятия будет предоставляться от 2-х до 5-ти минут на решение практической задачи. По истечению отведенного времени преподаватель решает задание на экране.

# 3 Теория вероятностей и мат. статистика

## 1 Основы теории графов и комбинаторика

- принцип Дирихле;
  - перестановки;
  - размещения;
  - сочетания;
  - задачи на сочетания;
  - «геометрические» задачи;
  - бином Ньютона;
  - определение графов и их применения.
- 

## 2 Дискретные случайные величины

- распределения случайных величин;
- дискретные и непрерывные распределения;
- функция распределения и её свойства;
- распределение Бернулли;
- риномиальное распределение;
- равномерное распределение;
- экспоненциальное распределение;
- распределение Лапласа;
- распределение Коши.

### Домашние задания

- 1 Вычислить мат.ожидание, написав программный код в Python.
- 

## 3 Непрерывные случайные величины

- распределения случайных величин;
  - дискретные и непрерывные распределения;
  - функция распределения и её свойства;
  - распределение Бернулли;
  - риномиальное распределение;
  - равномерное распределение;
  - экспоненциальное распределение;
  - распределение Лапласа;
  - распределение Коши.
-

## 4 Теоремы

- случайные последовательности и сходимость;
- теорема Пуассона для схемы Бернулли;
- закон больших чисел (Чебышева, Бернулли, Хинчина);
- ЦПТ Локальная;
- ЦПТ Линдеберга;
- ЦПТ Ляпунова.

### Домашние задания

- 1 Задачи на предыдущие темы курса: максимизировать функцию. Отнормировать относительно среднего. Вычислить количество возможных повторений.
- 

## 5 Точечное и интервальное оценивание

- точечные оценки и их свойства, метод моментов;
  - метод максимального правдоподобия;
  - асимптотическая нормальность оценок;
  - доверительные интервалы;
  - принцип построения доверительных интервалов.
- 

## 6 Регрессии

- линейные, многомерные, логистические регрессии;
- МНК, ММП, ММ.

### Домашние задания

- 1 Построить линейную регрессию в Python.
- 

## 7 Метод главных компонент

- определение главных компонент, их вычисление;
  - применение в задачах регрессионного анализа.
-

8 **Моделирование случайных величин** — условная функция распределения;  
— условная плотность распределения;  
— корреляционная зависимость.

---

9 **Моделирование случайных величин** метод Монте-Карло.  
Домашние задания

1 Провести Монте-Карло симуляцию для вычисления числа  $\pi$ .

---

10 **MidTerm** Проверочная работа. В ходе занятия будет предоставляться от 2-х до 5-ти минут на решение практической задачи. По истечению отведенного времени преподаватель решает задание на экране.

# 4 Проектная работа

- |   |   |   |
|---|---|---|
| 1 | <b>Вводное занятие по проекту</b>             | определение тем и заданий для финального проекта.<br><br>Домашние задания<br><br>1 Проектная работа |
| 2 | <b>АБ-тестирование в теории и на практике</b> | обсуждение кейса из практики.   |
| 3 | <b>Коллоборативная функция</b>                | обсуждение кейса из практики.   |
| 4 | <b>Консультации по проекту</b>                |   |
| 5 | <b>Защита проекта</b>                         | обсуждение результатов учащихся с выполнением проекта.  |