

Математика для Data Science. Продвинутый курс

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

Длительность курса: 132 академических часа

1 Линейная алгебра

- | | | |
|---|---|---|
| 1 | Введение 1. Математика в DataScience | понимание, что быть успешным DataScientist без знания математики, а главное без ее понимания, невозможно.
План и структура этого курса.
Знакомство с преподавателем |
|---|---|---|
-

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

- Базовые термины матанализа (предел, непрерывность функции, дифференциал)
- Базовые термины линейной алгебры (вектор, матрица, ее виды, ранг, определитель)
- Базовые термины теории вероятности (определение вероятности, мат.ожидание, дисперсия)
- Установка Python, предоставление рекомендаций по интерпретатору
- Вычисление базовых задач на Python с помощью стандартных пакетов

Домашние задания

- 1 Основы линейной алгебры, мат.анализа и теории вероятности

Цель: Цель данной домашней работы - развить практические навыки, полученных в ходе первого и второго уроков.

Рекомендуется сначала пройти урок 1 и 2.

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 **Геометрическая
интерпретация в
линейной
алгебре**

- Геометрическая интерпретация матричных преобразований
 - Правило Крамера
 - Знакоопределенность матрицы. Матрица Маркова. Жорданова форма
-

5 Матричные разложения

- Разложение SVD и ALS
- Неотрицательные разложения
- Заполнения пропусков в матрицах

Домашние задания

1 Разложить матрицу в SVD.

Цель: Выполнить сингулярное разложение в python и записать получившуюся матрицу преобразования для следующих матриц:

$$1) A = \begin{bmatrix} 3 & 7 \\ 5 & 2 \end{bmatrix}$$

$$2) B = \begin{bmatrix} 0 & 1 & 3 \\ 2 & 3 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$3) C = \begin{bmatrix} 5 & 7 & 6 & 4 & 2 \\ 1 & 3 & 4 & 1 & 3 \\ 3 & 4 & 5 & 4 & 1 \end{bmatrix}$$

1. Установить набор библиотек Anaconda for Python 3.7: <https://www.anaconda.com/distribution/>
2. Запустить графическую среду Ipython Notebook (написав в терминале `ipython notebook`, либо используя графический интерфейс Anaconda)
3. Отобразить код задания в Ipython Notebook
4. Прислать ссылку на репозиторий, в котором будет выложен Ipython Notebook с решением задачи

6 Матричные производные

- Матричные производные
 - Дифференциальные уравнения в матрицах
-

- 7 **Применение линейной алгебры в Data Science.** классификация данных с SVM и Logistic Regression
- Домашние задания
- 1 Повторить вычисления с занятия в Python на других данных.
-

- 8 **Применение линейной алгебры в Machine Learning** обработка изображений и линейная алгебра

- 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
- ММП

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 **Применение
Мат.анализа в ML**

- Покоординатный спуск
- Градиентный спуск
- Адаптивные варианты градиентного спуска
- Ньютоновские методы, BFGS

Домашние задания

1 Поиск экстремума с
Python.

9 **Применение
Мат.анализа в ML**

— Линейная регрессия и разные подходы к оптимизации (градиентный спуск, bounds, итд.)
— Нелинейная регрессия (реализация на нейронных сетях) и разные подходы к оптимизации

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

3 Теория вероятностей

1 Комбинаторика и Основы теории вероятностей

- Принцип Дирихле. Перестановки. Размещения. Сочетания.
 - Опыт и его исходы. Пространство элементарных событий. Вероятность события.
 - Независимость событий. Условная вероятность. Формулы сложения и умножения.
 - Формулы полной вероятности и Байеса
-

2 Случайные величины

- Случайная величина
- Дискретные и непрерывные случайные величины
- Закон распределения случайной величины и способы его описания
- Дискретные и непрерывные распределения
- Функция распределения и её свойства
- Распределение Бернулли
- Биномиальное распределение
- Моделирование на Python дискретное распределение (для задачи МонтеКарло)

Домашние задания

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

3 Непрерывные случайные величины

- Основные законы распределения и их физический смысл: биномиальное, пуассоновское, экспоненциальное, равномерное, гауссовское
 - Компьютерное моделирование различных распределений
-

- 4 **Теоремы**
- Случайные последовательности и сходимости
 - Теорема Пуассона для схемы Бернулли
 - Закон больших чисел (Чебышева, Бернулли, Хинчина)
 - ЦПТ Локальная, ЦПТ Линдеберга, ЦПТ Ляпунова
 - Точечные оценки и их свойства

Домашние задания

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

-
- 5 **Точечное и интервальное оценивание**
- Асимптотическая нормальность оценок
 - Доверительные интервалы. Принцип построения доверительных интервалов

-
- 6 **Проверка гипотез**
- Проверка статистических гипотез. Формулировка гипотез.
 - Проверка гипотезы о законе распределения выборки. Критерий согласия Пирсона.
 - Проверка гипотезы о независимости двух номинальных признаков. Критерий хи-квадрат

-
- 7 **Проверка гипотез. Часть 2**
- Ошибки I и II рода, уровень значимости. Статистический критерий. Построение доверительной и критической областей. P-value
 - Проверка гипотезы о вероятности успеха в схеме Бернулли. Биномиальный критерий
-

8 **Виды зависимостей** — Виды зависимостей случайных величин: функциональная, причинно-следственная, статистическая, корреляционная. Различия и связь между ними.
— Условные распределения

9 **Регрессии** — Линейные, многомерные, логистические регрессии
— МНК, ММП, ММ

Домашние задания

1 Построить линейную регрессию в Python.

10 **Метод главных компонент** — Определение главных компонент, их вычисление
— Понижение размерности и отбор признаков
— Применение в задачах регрессионного анализа

11 **Моделирование случайных величин** — Random Walks, практические методы
— Вероятностный классификатор: Logistic Regression, Random Forest, Gradient Boosting
— Вероятностные Нейронные Сети

12 **Моделирование случайных величин. Часть 2** — Random Walks, практические методы
— Вероятностный классификатор: Logistic Regression, Random Forest, Gradient Boosting
— Вероятностные Нейронные Сети

Домашние задания

1 Провести Монте-Карло симуляцию для вычисления числа Пи.

13 **MidTerm**

- Проверочная работа.
- Обсуждение тем проектной работы

4 Проектная работа

- | | | |
|---|----------------------------------|--|
| 1 | Консультация по проекту | обсуждение кейса из практики

Домашние задания

1 Проектная работа |
| 2 | Консультация по проекту | обсуждение кейса из практики |
| 3 | Защита дипломного проекта | обсуждение кейса из практики |