Skillbox

Основы Python (2019)

* Введение в Python
* Основы Python: установка PyCharm
* Основы Python: базовые структуры данных
* Основы Python: циклы и условия
* Основы Python: функции
* Основы Python: классы и объекты

Перегрузка операторов

Различие атрибутов класса и экземпляра

Пространство имен класса

Наследование классов

Переопределение свойств и методов родителя

Метод super

Множественное наследование

* Основы Python: исключения

Обработка исключений

Исключение в контексном менеджере

Класс Warning

* Базовые коллекции: Строки
* Базовые коллекции: словари и множества
* Базовые коллекции: Кортежи
* Инкапсуляция (соктрие данных)
* Наследование
* Полиморфизм
* Работа с файлами
* Функции — Рекурсия

Аналитика. Начальный уровень

* Библиотека NumPy: методы анализа массивов
* Библиотека pandas: индексация и выбор данных

Основные структуры.

Индексы и метод .iloc.

Индексы и метод .loc.

Характеристики датафрейма pandas.

Операции над датафреймами.

Применение функций и метод .apply ( ).

Группировка и агрегирование.

Сортировка и упорядочивание.

Работа с NaN.

Иерархическое индексирован

* Основы визуализации данных с помощью Matplotlib

Введение в визуализацию.

Знакомство с Matplotlib.

Детали графика (подписи, заголовки, легенды и прочее).

Виды данных, визуализация численных данных.

Визуализация категориальных данных.

Частые ошибки визуализации.

Жизненный цикл графика matplotlib.

Визуализация в цикле, несколько графиков на одной координатной сетке.

Несколько графиков на одном полотне.

Стилизация графиков. Часть 1.

Стилизация графиков. Часть 2.

* Визуализация с Seaborn

Что такое seaborn.

Построение основных графиков в seaborn.

Построение составных графиков.

Стилизация графиков seaborn

* Чтение и запись данных: CSV, XLSX

Что такое данные и какие они бывают.

Чтение файлов в формате csv.

Запись файлов в формате csv.

Форматирование дат.

Чтение и запись файлов в формате XLSX.

* Основы SQL

Введение в SQL.

Структура БД, запрос DESCRIBE.

Выбор и фильтрация данных, запрос SELECT.

Функции и выражения, агрегация данных.

Отношения и соединение таблиц: JOIN.

Группировка.

Модификация данных.

Подзапросы.

Объявление и изменение структуры данных

* Чтение и запись данных: JSON, MongoDB

Чтение данных JSON из веб-сервисов.

Чтение и запись файлов в формате JSON.

Чтение из баз данных с помощью pandas и sqlalchemy.

Работа с большими данными — MongoDB

* Работа со строками

Введение.

Стандартные приёмы работы с текстом.

Регулярные выражения Python.

Статистики текста.

# Основы статистики и теории вероятностей

Распределение вероятности.

Оценки распределения случайной величины.

Анализ распределения случайной величины — симметричные данные.

Проклятие асимметричности, бимодальность, или где чаще всего ошибаются аналитики. домашнее задание 11 1 уроков

Анализ распределения случайной величины — асимметричные данные.

Корреляционный анализ.

Ищем взаимосвязи в данных.

Доверительные интервалы.

Построение доверительной оценки.

Аномалии в данных.

Поиск аномалий в датасете

Аксиоматический подход – базирование решений на реальной выборке

Распределиние может быть

* Дескретное (результат футбольного матча) (Пуассон, Бернулли, Биномиальное)
* Непрерывное (цена доллара) (Гаусов закон, Гамма)

Плотноть распределения случайной величины – вероятность попадания случайной величины на участок х + дельта х

Математическое ожидание или среднее значение случайной величины m(x) expected value

Дисперсия – показывает рассеяние значений вокруг величины

Среднеквадратичное отклонение СИГМа

Мода – самое часто встречаемое значение случайной величины (самая частая покупка)

Квантили – семейство оценок.

Квартиль – вероятность 0.25 0.5 0.75 4 равных промежутка. 0.5 – медиана

Интерквартильный размах – полезен для оценка разброса 0.75 – 0.25 квартили

# Основы математики для Data Science

* Базовые математические объекты и SymPy. Дроби и преобразования

Основные операции с дробями.

Раскрытие скобок.

Приведение подобных слагаемых.

SymPy. Работа с математическими формулами в Python

* Базовые математические объекты и SymPy. Функции и дополнительные объекты

Степени и их свойства.

Корни и их свойства.

Логарифмы и их свойства.

Дополнительные объекты и обозначения

* Функции одной переменной, их свойства и графики

Понятие функции. Построение графиков функций с помощью SymPy.

Виды и свойства функций.

Элементарные функции и их свойства.

Исследование параболы с помощью SymPy.

Дополнительный функционал SymPy для исследования функций.

* Интерполяция и полиномы

Полиномы и интерполяция.

Свойства коэффициентов квадратичной функции.

Свойства коэффициентов кубической функции.

Аналитический способ нахождения коэффициентов полиномов.

Недостатки интерполяции.

Интерполяция — это метод нахождения неизвестных промежуточных значений некоторой функции по имеющемуся дискретному набору ее известных

* Аппроксимация и преобразования функций

Функция потерь.

Преобразования функций. Сдвиги вправо и влево.

Преобразования функций. Растяжения и сжатия.

ML как решение задачи аппроксимации.

* Аппроксимация и производные

Понятие производной.

Минимум MSE и техники вычисления производных (часть 1).

Минимум MSE и техники вычисления производных (часть 2).

Сигмоида и ещё несколько правил вычисления производных.

* Функции нескольких переменных, их свойства и графики

Определение функции нескольких переменных.

Графики функции нескольких переменных.

Нули функции нескольких переменных.

Декартова система координат и гиперплоскость.

* Частные производные функции нескольких переменных

Задача аппроксимации функцией многих переменных.

Техника нахождения частных производных.

Частные производные и минимизация MSE (часть 1).

Частные производные и минимизация MSE (часть 2).

Нахождение производных с помощью SymPy.

* Векторы и матрицы. Градиент

Векторы как описания объектов и их признаков.

Скалярное произведение векторов и линейная модель.

Геометрическая интерпретация вектора.

Вектор функций и градиент

Гралиент зависит от частных производных и получается вектор

* Линейная регрессия и системы линейных уравнений

Линейная регрессия (Linear regression) — модель зависимости переменной x от одной или нескольких других переменных

Детерминант.

Обратная матрица.

Регуляризация.

Регрессия vs классификация.

Степень уверенности: расстояние от прямой до точки.

* Задача аппроксимации как матричное уравнение

Решение задачи аппроксимации с помощью матриц.

Детерминант и нахождение обратных матриц (часть 1).

Детерминант и нахождение обратных матриц (часть 2).

Псевдообратная матрица

# Машинное обучение. Начальный уровень

* Основные концепции Machine Learning

Введение в ML.

Кейсы применения ML в реальной жизни.

Обучение с учителем. Регрессия.

Обучение с учителем. Классификация.

Обучение без учителя. Кластеризация.

Обучение с подкреплением. Нейронные сети и ансамбли