

Практическая работа №8

Метод k -средних.

Оглавление

Цель работы.....	1
Задачи работы.....	1
Перечень обеспечивающих средств	1
Общие теоретические сведения	1
Метод k -средних оптимизирует (ищет минимум) величину инерции — сумму квадратов расстояний от всех элементов набора данных до центров построенных кластеров.Задание 1	1
Требования к предоставлению результатов	3
Литература.....	3

Цель работы

Практическое использование метода k -средних для задачи кластеризации.

Задачи работы

1. Научиться использовать метод k -средних.
2. Научиться анализировать качество кластеризации методов k -средних.

Перечень обеспечивающих средств

1. ПК.
2. Учебно-методическая литература.
3. Задания для самостоятельного выполнения.

Общие теоретические сведения

Задача кластеризации

Задача кластеризации — это пример задачи с обучением без учителя. В данной задаче рассматривается набор данных, о которых нет дополнительной информации о принадлежности к тем или иным классам.

Тем не менее, есть методы, позволяющие на основании анализа данных разделять их на кластеры по различным признакам.

Метод k -средних и анализ качества кластеризации

Метод k -средних используется для кластеризации на основе расстояния между отдельными элементами набора данных.

Можно использовать любое определение расстояния. В случае, если набор данных состоит из чисел или наборов (векторов) чисел, логично использовать простое Евклидово расстояние — квадратный корень из суммы квадратов разностей отдельных координат.

Метод k -средних оптимизирует (ищет минимум) величину инерции — сумму

квадратов расстояний от всех элементов набора данных до центров построенных кластеров.

Задание

Пояснение

- Для сохранения результатов данной работы вам понадобится два файла: doc/docx – для текста и ipynb – для кода. Назовите их одинаково: «*Фамилия* – задание 8».

Часть 1

- Обновите свой репозиторий, созданный в практической работе №1, из оригинального репозитория:
https://github.com/mosalov/Notebook_For_AI_Main.

Часть 2

- Откройте свой репозиторий в Binder (<https://mybinder.org/>).
- Откройте файл «task8.ipynb».
- Выполните представленный в файле код, чтобы понять, как применять метод k-средних с использованием библиотеки sklearn.
- Выполните инструкции, указанные в комментариях во второй части файла, начиная с «Загрузите файл "worldcities.csv" в DataFrame.»
- Сохраните код в ipynb-файле. Необходимые пояснения опишите в своём docx/doc-файле.

Требования к предоставлению результатов

Оба файла (doc/docs и ipynb) загрузите в свой репозиторий, созданный в практическом задании №1 по пути: «Notebook_For_AI_Main/2021 Осенний семестр/Практическое задание 8/» и сделайте пул-реквест.

Литература

1. <http://datascientist.one/k-means-algorithm/>
2. <https://craftappmobile.com/кластеризация-методом-k-средних/>
3. <http://datareview.info/article/klasterizatsiya-s-pomoshhyu-metoda-k-srednih-na-python/>
4. <https://proglab.io/p/unsupervised-ml-with-python/>