

Домашнее задание 8

Авласов Владислав

Задание 3

Реализовал k-means. Никаких улучшений в алгоритм не вносил, поэтому первоначальные центры кластеров выбираются случайно. Из-за этого на некоторых запусках точность довольно низкая, скорее всего из-за того, что кластеры выбрались очень близко и для хорошей сходимости пришлось бы очень долго ждать, больше 100 итераций, которые алгоритм работает.

Тем не менее, запустив несколько раз, тем самым избежав неудачных выборов центров, можно заметить, что точность кластеризации растёт при росте параметра s и достигает 1. Что в принципе логично, т.к. наши множества отдаляются друг от друга.

Не придумал красивый способ получения соответствия лейблов и в итоге просто перебирал все варианты в поиске лучшего, с максимальной точностью. Графики снизу. Не знаю, почему латех не вставляет их между текстами.

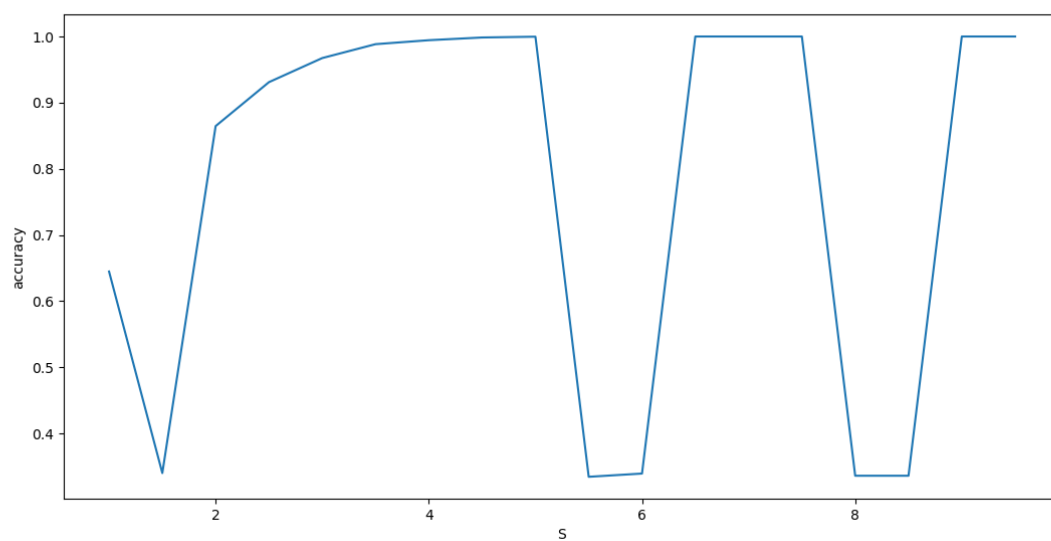


Рис. 1: Неудачные выборы начальных центров сказываются на точности, но можно заметить, что при росте S точность удачных запусков растёт.

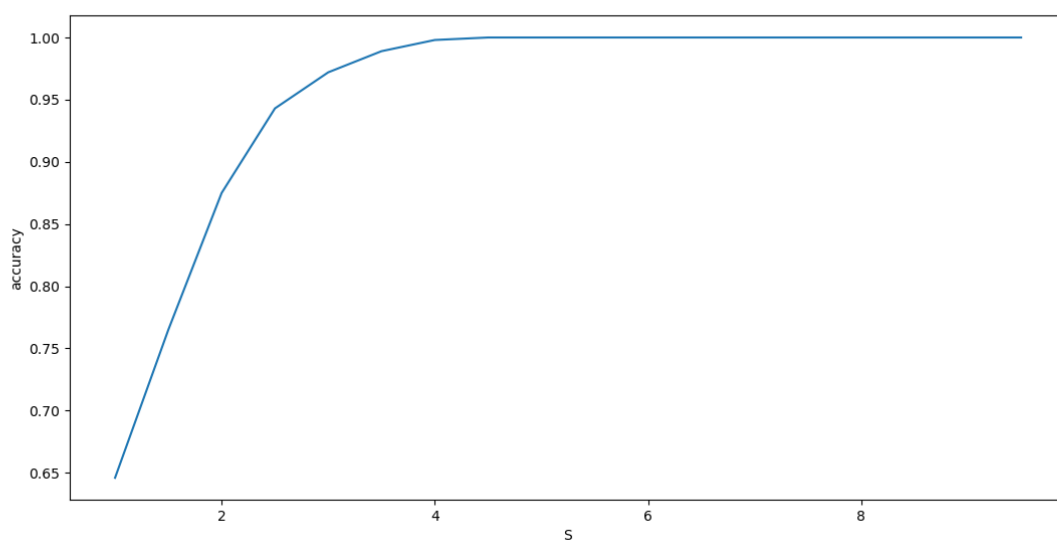


Рис. 2: При запуске на каждом значении s по 10 раз мы откинули самые неудачные запуски, поэтому график получился плавным и логичным.