

Домашнее задание 7

Авласов Владислав

Задание 2

Код доступен в `task2_alternative.py`.

Датасет очень большой, поэтому пришлось неожиданно долго ждать. В итоге я выбрал из середины датасета 10 тысяч объектов для тренировочной выборки и 2 тысячи объектов для тестовой.

Точность на несжатом датасете = 0.929. Сжатие даже помогло нам убрать некоторые лишние шумы и выделить самые важные характеристики!

Используя PCA из библиотеки *sklearn.decomposition*, запускаем наш алгоритм knn с $k=10$ на размерностях **[5, 10, 20, 50, 100, 150, 200, 250, 300, 350, 400, 450]**. Точности оказались одинаковые для сжатых и восстановленных версий изображений. Поэтому в коде закомментированы строчки с запуском алгоритма на восстановленных версиях, ибо это невероятно долго. Максимальная точность достигается уже при сжатии до размерности $n = 50$ и равна она = 0.9645, и дальше не растёт, где-то даже начинает падать. График:

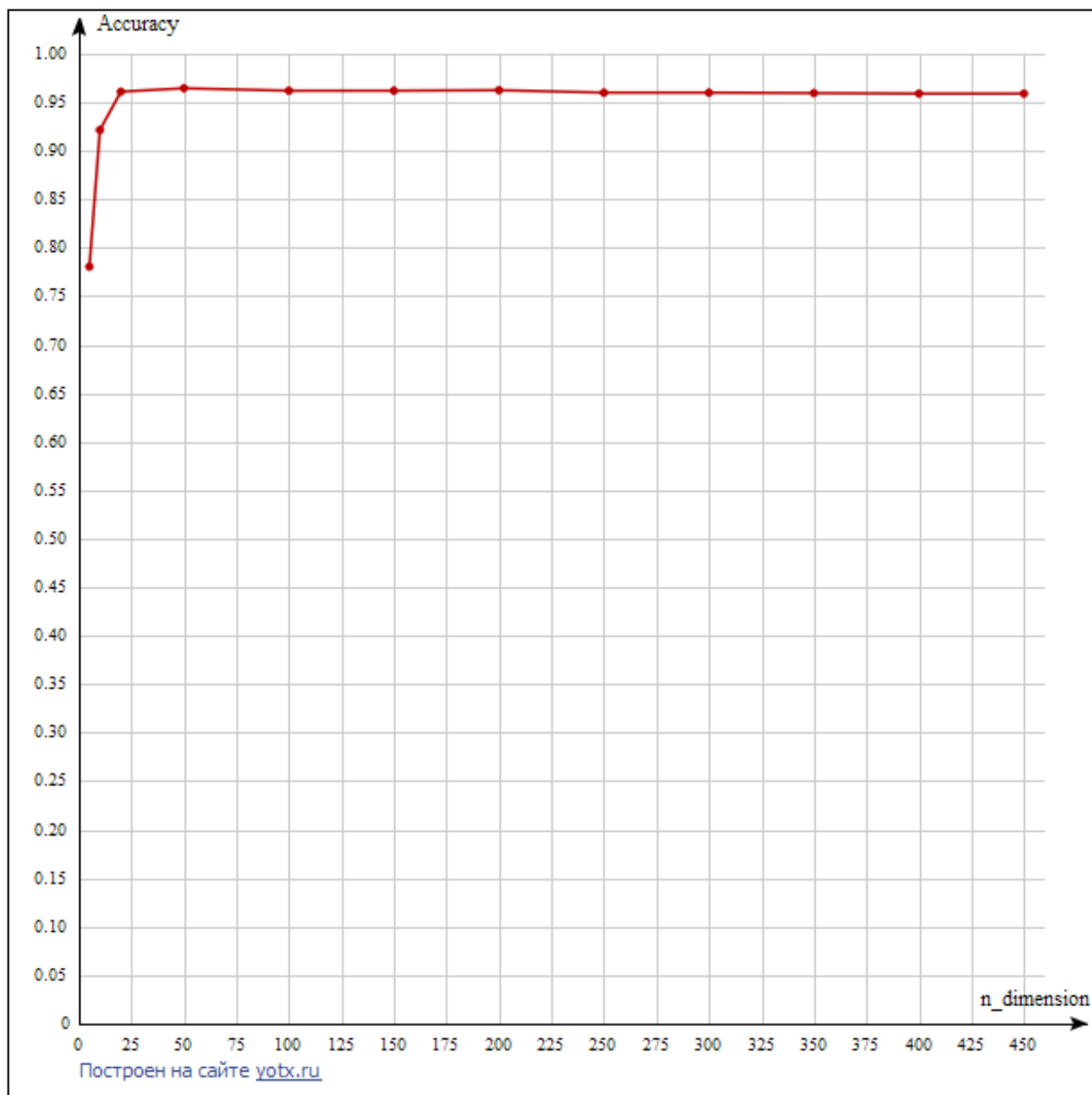


Рис. 1: Лучший результат при $n=50$, $\text{acc} = 0.9645$