

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский университет ИТМО»
(Университет ИТМО)

Факультет систем управления и робототехники

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №6
по дисциплине
«Практическая линейная алгебра»

по теме:
SINGULAR VALUE DECOMPOSITION

Студент:
Группа № R3435

Зыкин Л. В.

Предподаватель:
техник, ассистент

Догодин Е. В.

Санкт-Петербург
2025

Задание 1. Сжатие изображений с помощью SVD

В качестве исходного изображения использована картинка из набора по предмету, переведённая в оттенки серого и представлена в виде матрицы интенсивностей $A \in \mathbb{R}^{m \times n}$.

Было выполнено SVD-разложение $A = U \Sigma V^T$, после чего построены укороченные аппроксимации $A_k = U_{[:,1:k]} \Sigma_{1:k,1:k} V_{[1:k,:]}$ для ряда значений k . Ниже приведены исходное изображение и сетка реконструкций для девяти значений k , а также графики ошибки восстановления (MSE) и относительно-го объёма хранения.

Численно степень сжатия оценивалась как

$$\text{ratio}(k) = \frac{mk + k + kn}{mn},$$

где числитель — количество значений для хранения укороченного разложения U_k, Σ_k, V_k . Для представленных значений k график на рисунке 3 показывает компромисс между качеством (MSE) и объёмом хранения.

Исходное изображение



Рисунок 1 — Исходное изображение в оттенках серого

Анализ результатов

- При малых k (1–3) картинка сильно теряет детали и становится расплывчатой; она уже распознаваема, но качество невысокое.
- При средних k видны основные контуры и крупные детали; качество приемлемое при заметном снижении объёма хранения.

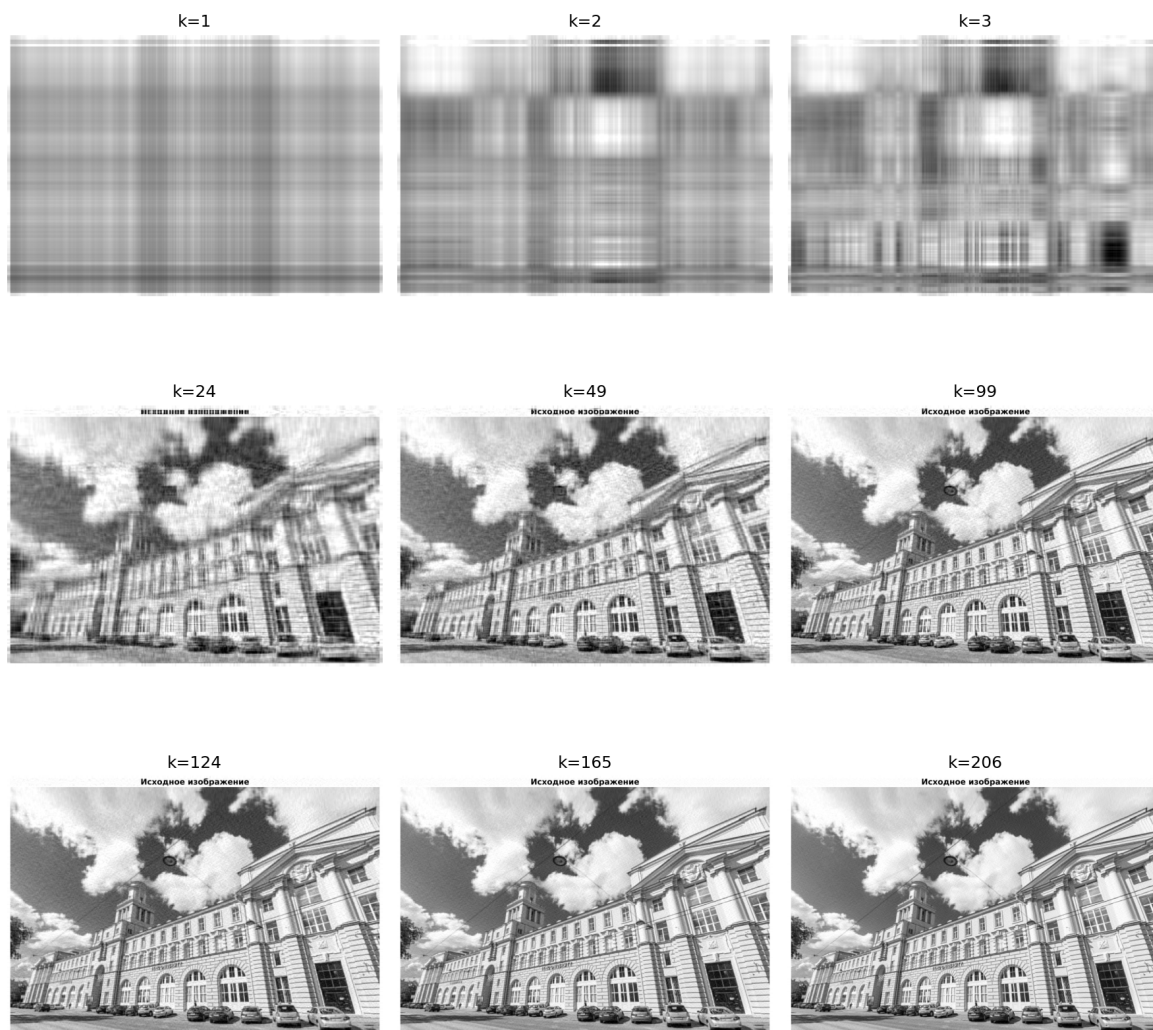


Рисунок 2 — Реконструкции при различных k (9 вариантов)

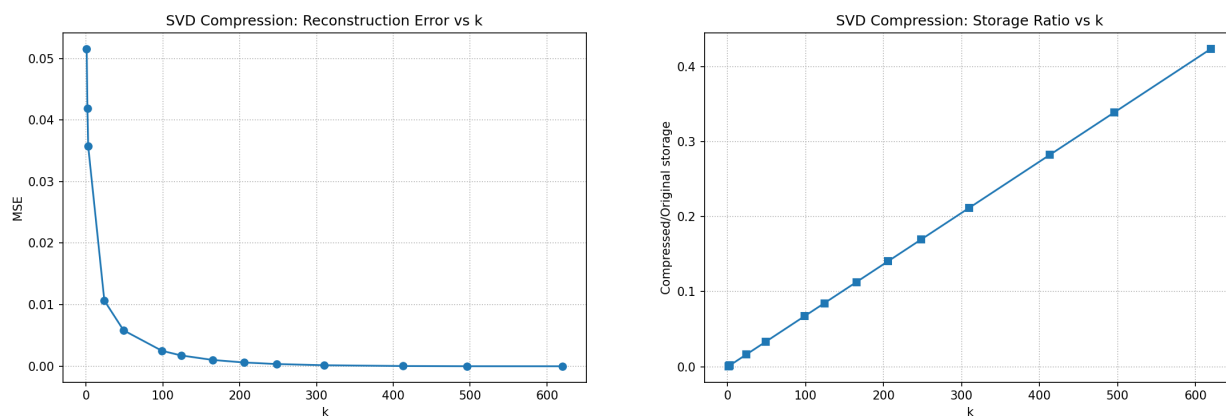


Рисунок 3 — Зависимость ошибки (MSE) и относительного объёма хранения от k

- При больших k визуально близко к исходному, но выигрыш по памяти меньше.

Задание 2. Latent Semantic Analysis (LSA)

Для набора текстовых документов был выполнен предобработчик: удаление пунктуации и цифр, приведение к нижнему регистру, токенизация, лемматизация (ru morphology2) и исключение стоп-слов (встроенный список). По очищенной коллекции построена терм-документная матрица (TF-IDF-подобное взвешивание), затем выполнено SVD-разложение.

В качестве корпуса использован файл PLA/lab6/python/data12.txt (22 коротких документа).

Ниже приведены: спектр сингулярных чисел, а также визуализации двух тем — топ-5 слов в каждой теме и облака слов.

Дополнительно построено «облако слов» по всему корпусу (без разбиения на темы).

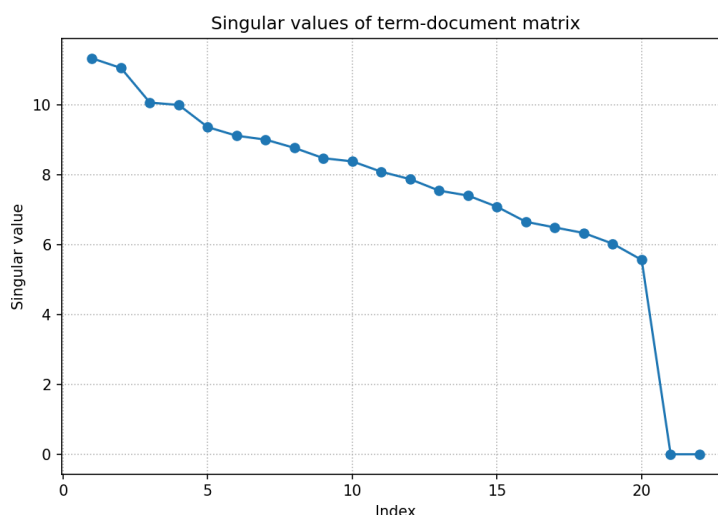


Рисунок 4 — Сингулярные числа терм-документной матрицы

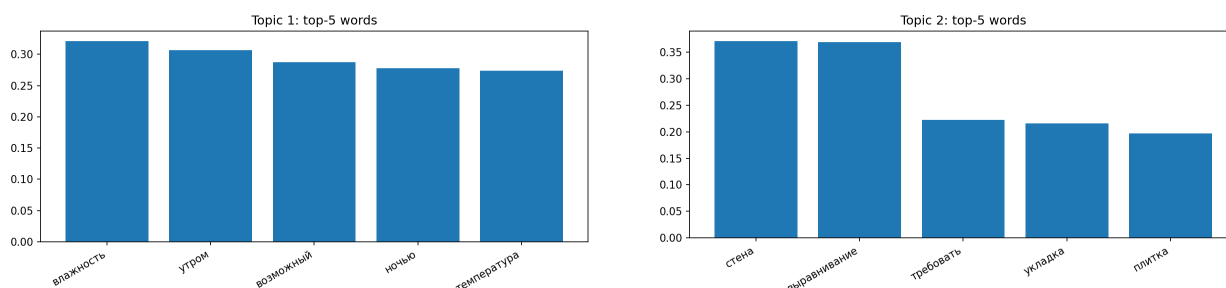


Рисунок 5 — Топ-5 слов по абсолютному вкладу в двух ведущих темах

Выводы по проделанной работе

- По задаче SVD-сжатия: при малых k изображение заметно теряет детали; при средних k качество уже приемлемое при разумном уменьшении объёма хранения; при больших k картинка близка к исходной, но выигрыш по памяти снижается. Характерная зависимость MSE и доли хранения от k подтверждает ожидаемый компромисс «качество—сжатие».
- По задаче LSA: базовая предобработка (pymorphy2+стоп-слова) и SVD выделили две устойчивые темы — погода/прогноз и ремонт/отделка; связанные документы определяются по весам правых сингулярных векторов. Убывание сингулярных чисел указывает на низкоранговую структуру.

Приложение

Текстовый корпус для LSA

1. Днём температура воздуха поднимется до +20 градусов, а ночью сохранится влажность.
2. Перед поклейкой обоев важно очистить стены от старого покрытия.
3. На юге страны прогнозируются сильные ветры и проливные дожди.
4. Шпаклёвка позволяет выровнять стены перед поклейкой обоев.
5. Ветер с порывами до 20 м/с принесёт ливневые осадки и облачность.
6. Для укладки плитки требуется подготовить поверхность и выбрать клей.
7. Температура остаётся низкой, прогнозируются осадки и облачность.
8. Гипсокартон используют для выравнивания стен и монтажа перегородок.
9. Пасмурная погода сохранится, местами возможен сильный ветер.
10. Укладка ламината требует ровного основания и подложки для изоляции.
11. Завтра утром ожидается густой туман и высокая влажность.
12. Перед установкой потолочных панелей важно снять старую штукатурку.

13. Ливневые дожди продолжатся до конца недели, местами будет туман.
14. Штукатурка применяется для выравнивания стен перед окрашиванием или укладкой плитки.
15. Температура воздуха утром будет низкой, а днём возможны прояснения.
16. Кафельную плитку кладут на специальный клей, тщательно проверяя уровень.
17. Синоптики предупреждают о тумане и гололёде на дорогах ночью.
18. Замена труб в ванной требует точного расчёта и аккуратного выравнивания стен.
19. Ночью возможны заморозки, а утром снова прогнозируют высокую влажность.
20. Для ремонта пола используют цементную смесь для выравнивания и покрытия ламинатом.