Звіт:

<https://machinelearningmastery.com/singular-value-decomposition-for-machine-learning/> --- svd

<https://www.geeksforgeeks.org/how-to-compute-the-eigenvalues-and-right-eigenvectors-of-a-given-square-array-using-numpy/> ---- eigenvectors and eigenvalues

<https://medium.com/@ritik_gupta/how-singular-value-decomposition-svd-is-used-in-recommendation-systems-clearly-explained-201b24e175db>

<https://jaketae.github.io/study/svd/>

Зображення, що містить текст, схема, ряд, знімок екрана

Автоматично згенерований опис

Мета використання SVD для факторизації матриці «користувач-елемент» полягає в тому, щоб знайти дві нові матриці, одна з яких представляє користувачів, а інша — елементи, які, перемножені разом, якомога точніше наближають вихідну матрицю. Ці нові матриці називаються прихованими представленнями, і вони містять важливу інформацію про користувачів і елементи, які можна використовувати для надання рекомендацій.

1. Які основні етапи включає SVD розклад і як цей метод можна застосувати до вирішення задачі підбору рекомендацій для певного користувача?

**Етапи SVD розкладу:** # A=Um\*m(otrhogonal- rotate) \* Em\*n(diagonal - stretch) V\_transponedn\*n(otrhogonal - rotate)

1. **Формування матриці:**
   * Створення матриці R, де рядки відповідають користувачам, а стовпці — об'єктам (наприклад, фільмам), заповненої оцінками.
2. **Нормалізація даних:**
   * Віднімання середнього значення оцінок для кожного користувача, щоб отримати нормалізовану матрицю RR.
3. **Виконання SVD:**
   * Розкладання нормалізованої матриці RRR на три матриці: UUU, Σ\SigmaΣ і VTV^TVT: R=UΣVTR = U \Sigma V^TR=UΣVT де:
     + UUU — матриця лівих сингулярних векторів (розміром m×km \times km×k),
     + Σ\SigmaΣ — діагональна матриця сингулярних значень (розміром k×kk \times kk×k),
     + VTV^TVT — матриця правих сингулярних векторів (розміром k×nk \times nk×n).
4. **Відновлення матриці:**
   * Відновлення нормалізованої матриці RRR з використанням зменшеної кількості сингулярних значень: Rapprox ​=UΣVT

**Прогнозування:**

* Відновлення оригінальної матриці оцінок додаванням середніх значень, віднятих на етапі нормалізації, щоб отримати прогнозовані оцінки для всіх користувачів і об'єктів.

2.В яких сферах застосовується SVD?

**Застосування до задачі підбору рекомендацій:**

* **Формування матриці оцінок:** Зібрати дані про оцінки користувачів.
* **Нормалізація:** Відняти середнє значення оцінок для кожного користувача.
* **Розклад SVD:** Розкласти матрицю оцінок на три матриці.
* **Відновлення матриці:** Відновити нормалізовану матрицю і додати середні значення.
* **Прогнозування:** Використовувати відновлену матрицю для прогнозування оцінок для фільмів, які користувач ще не оцінив.
* **Рекомендації:** Вибрати топ-10 фільмів з найвищими прогнозованими оцінками для конкретного користувача.
* Компресія зображень: SVD використовується для компресії зображень шляхом зниження розмірності даних зображення.
* Обробка природної мови: SVD використовується для зниження розмірності текстових даних та виконання тематичного моделювання.
* Виправлення даних: SVD використовується для виправлення пропущених значень в даних.
* Виявлення аномалій: SVD використовується для виявлення аномалій в даних.

**Сфери застосування SVD**

* **Рекомендаційні системи:** Наприклад, для рекомендацій фільмів, музики, товарів на основі оцінок користувачів.
* **Обробка зображень:** Зменшення розмірності зображень для зберігання або передачі.
* **Текстова аналітика:** Аналіз текстових даних, виділення основних тем (латентно-семантичний аналіз).
* **Біоінформатика:** Аналіз геномних даних для виявлення важливих біологічних маркерів.
* **Розпізнавання голосу і зображень:** Для зменшення розмірності векторів ознак і покращення якості розпізнавання.

3.Як вибір параметра k у SVD розкладі впливає на результат?

**Вплив вибору параметра k у SVD розкладі**

* **Занадто малий k:**
  + Модель може втратити важливу інформацію, що призведе до низької якості прогнозів.
  + Зменшена матриця буде занадто простою і не зможе адекватно відображати складні взаємозв'язки в даних.
* **Занадто великий k:**
  + Модель може зберегти занадто багато шуму, що призведе до переобучення.
  + Збільшується обчислювальна складність і час виконання алгоритму.
* **Оптимальний k:**
  + Забезпечує баланс між збереженням важливої інформації та усуненням шуму.
  + Потрібно підбирати емпірично або за допомогою методів крос-валідації.

4.Які основні переваги та недоліки має SVD?

**Основні переваги та недоліки SVD**

**Переваги:**

* **Зменшення розмірності:** Дозволяє зменшити розмірність даних, зберігаючи найбільш важливу інформацію.
* **Шумозахист:** Допомагає зменшити вплив шуму на дані.
* **Універсальність:** Застосовується в багатьох сферах, де потрібне зменшення розмірності або виділення основних компонент.
* **Прогнозування:** Ефективний метод для прогнозування відсутніх даних в матрицях (наприклад, рекомендаційні системи).

**Недоліки:**

* **Обчислювальна складність:** SVD є обчислювально витратним процесом, особливо для великих матриць.
* **Чутливість до даних:** Результати можуть сильно змінюватися залежно від якості і розподілу даних.
* **Параметр k:** Вибір параметра k може бути складним і критичним для якості результатів.
* **Інтерпретація:** Сингулярні вектори можуть бути важко інтерпретувати в контексті реальних даних.