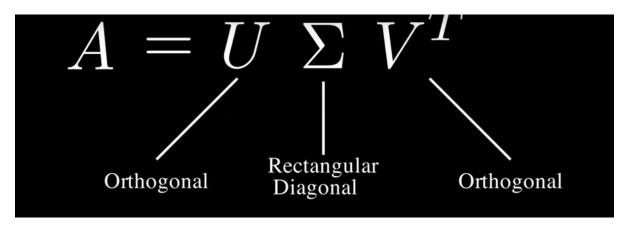
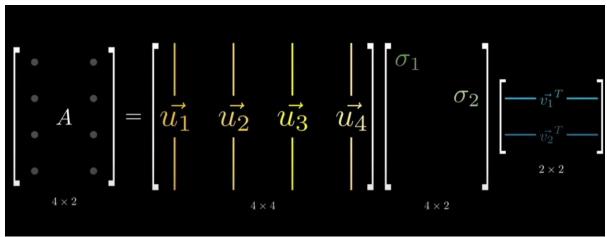
Які основні етапи включає SVD розклад і як цей метод можна застосувати до вирішення задачі підбору рекомендацій для певного користувача?





Якщо нам просто дана матриця А:

1) Обчислення U , Σ та VT

U – складається з власних векторів ААТ

 Σ — складається з квадратного кореня власних значень U розташованих на спадання

Діагональна матриця – розтягує кожну вісь

Ортогональна створює обертання

Як застосовувати до підбору рекомендацій:

- 1) Збір даних (збір рейтингу користувачів для фільмів чи музики)
- 2) Заповнення відсутніх значень нулем або середнім рейтингом користувача
- 3) Віднімаємо середні рейтинги кожного користувача від усіх його оцінок для отримання матриці відхилень Rdemeaned
- 4) Виконуємо сингулярний розклад відхилень

- U -матриця лівих сингулярних векторів (користувачі-фактори).
- Σ діагональна матриця сингулярних значень.
- VT матриця правих сингулярних векторів (фільми-фактори).
- 5) Зменшення розмірності
- 6) Реконструкція матриці
- 7) Обчислення прогнозованих рейтингів:
- 8) Генерація рекомендацій. Для кожного користувача визначити фільми, які він ще не бачив, і сортувати їх за прогнозованими рейтингами. Рекомендувати ті фільми, які мають найвищі прогнозовані рейтинги.

В яких сферах застосовується SVD?

SVD — це неймовірно потужний спосіб обробки даних

Застосовуємо для:

- 1) для побудови рекомендаційних систем, моделей і аналізу рейтингів(аналітика, економіка та фінаси)
- 2) Машинного навчання (зменшення розмірності)

Як вибір параметра k y SVD розкладі впливає на результат?

Чим менше значення k означає що ми залишаємо найбільш значущі сингулярні значення та вектори. Може призвести до більшої втрати інформації але зменшити розмірність. Також втрачає деталі та точність

Якщо значення k <u>більше</u>, то ми можемо зберегти більше інформації , забезпечує більше оригінальних даних

Які основні переваги та недоліки має SVD?

ПЕРЕВАГИ:

- 1) Зменшення розмірності
- 2)Використання в рекомендаційних системах

НЕДОЛІКИ:

- 1)Потрібно заповнювати пропущенні дані
- 2) Може бути втрата деталей при зменшенні розмірності

Зменшення розмірності означає що замість 5 записів у кожному рядку, як у вихідній матриці рейтингів A, тепер у нас ϵ 3 записи в рядку. Іншими словами, інформація про користувачів була стиснута в трьох вимірах. На відміну від того, де кожен стовпець відповідав якомусь фільму, тепер ми не знаємо що означають стовбці у тривімірному масиві. Це може бути якийсь жанр, актриса або будь-які приховані шаблони в наборі даних, про які ми не знаємо.

Додавання великої кількості нульових значень у матрицю знизить середній рейтинг для фільмів, що може вплинути на рекомендаційні алгоритми. В результаті, фільми, які насправді мають позитивні оцінки, можуть бути понижені в рейтингу.