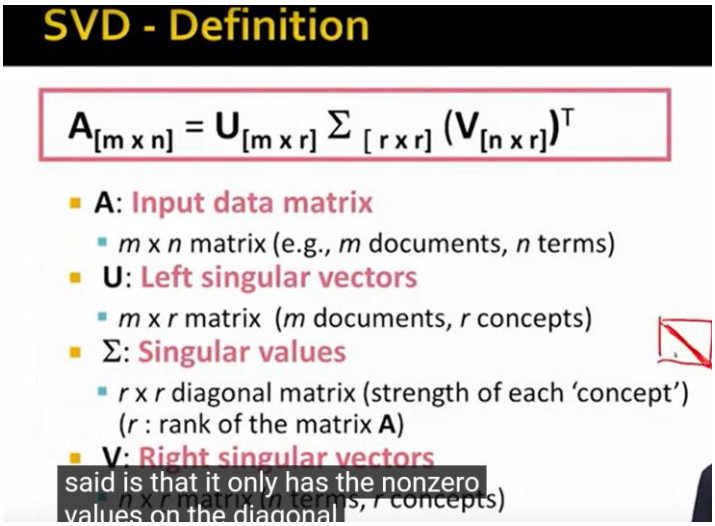
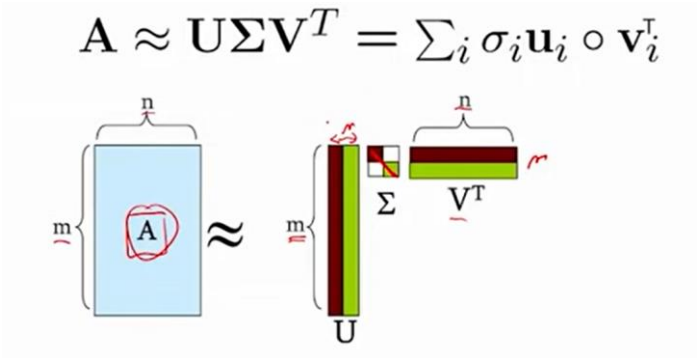
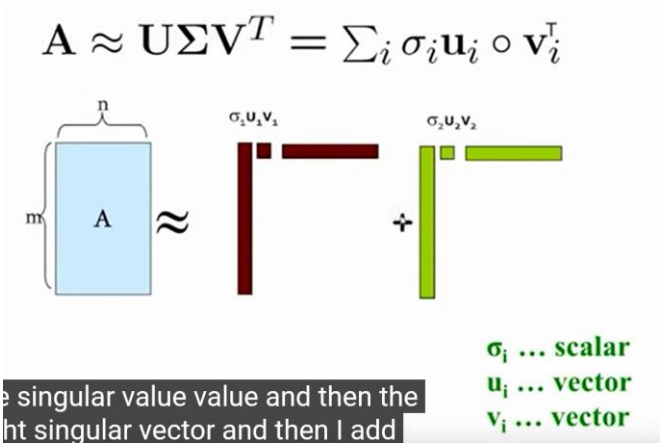
SVD



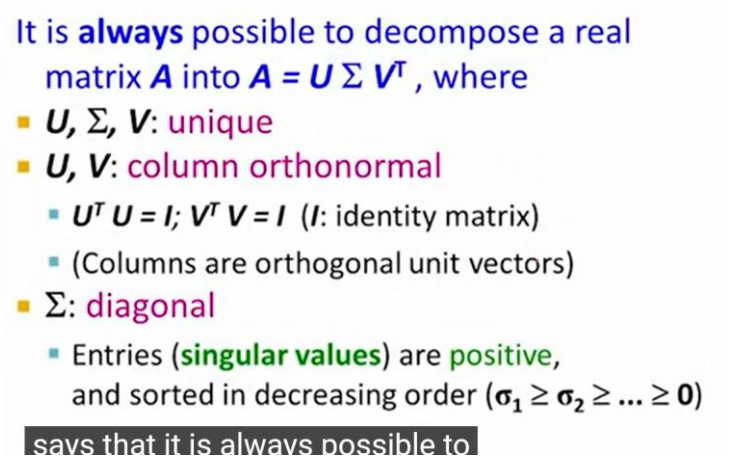
Dalam video tersebut svd merupakan salah satu dimensional reduction technique. Dalam svd input dinamakan data matriks yang terukur dalam n * n. dimana n * n merupakan m dan n, m sebagai baris dan n sebagai kolom. Dalam rumus tersebut A merupakan input data matriks; U merupakan left singular vector yaitu pengkalian m terhadap r, r diasumsikan dengan nilai yang kecil dan rank matriks; Σ sigma merupakan singular values dimana untuk penghitungan ini dilakukan secara diagonal terhadap nilai non zero values; V yaitu Right singular vector dimana penghitungan terhadap baris dengan nilai r.



A sebagai input, yang mana masukannya memiliki m sebagai baris dan n sebagai kolom. Dari masukan A akan di mempresentasi matriks A sebagai 3 produk matriks U, Σ dan V transpose. Gambar U Digambar dengan dengan tipis berdiri tetapi memiliki sedikit kolom dan banyak baris, Σ sigma Digambar dengan kotak kecil yang hanya memiliki beberapa elemen diagonal, serta V yang memiliki sedikit baris tetapi kolomnya banyak.



σ atau Σ merupakan scalar mempresentasi left singular vector sebagai outer produk vector yang berbeda. Kemudian akan dikalikan dengan U sebagai vector atau singular value kemudian V sebagai right singular value. Lalu ditambah dengan left singular, singular value, dan right singular value kedua. Hasil pengolahan tersebut didapatkan dari pengkalian terhadap m dan n dari U, V, dan Σ.



Matriks real A selalu dapat didekomposisi

menjadi A = UΣ VT, dimana

■ U, Σ, V: unik / atau berbeda

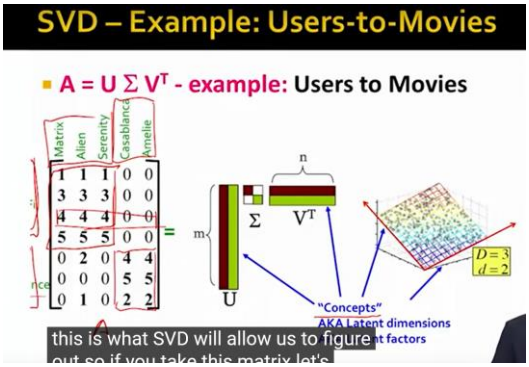
■ U, V: kolom ortonormal

UT U = 1; VT V = I (1: matriks identitas)

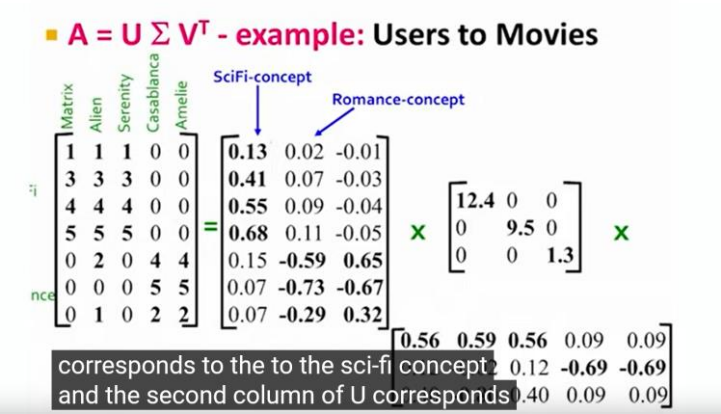
(Kolom adalah vektor satuan ortogonal)

■ Σ: diagonal

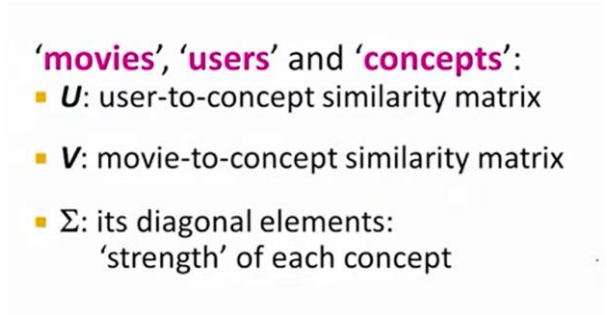
• Entri (nilai tunggal) adalah positif, dandiurutkan dalam urutan menurun (o, ≥ 6, ... ≥0)



Gambar tersebut terdapat kolom film dengan barisnya yaitu user, data matriks tersebut akan di kelompokan menjadi beberapa grup baris dan kolom dengan metode svd akan mempresentasekan.



Gambar tersebut merupakan contoh dari rumus diatas. Yang mana A = U Σ V. pada U nilai 0.13 merupakan nilai user penggemar terhadap scifi concept yang nilainya berkorespon dengan concept, begitu juga nilai -0.59 yang diminiliki oleh user penggemar romance concept. Pada Σ nilai 12.4 merupakan nilai singular values sigma, dalam matriks itu nilai 12.4 merupakan nilai yang kuat dan non zero positif dimiliki oleh user scifi concept. Pada nilai V merupakan ‘movie to concept’ similarity concept dimana nilai 0.56, 0.59, dan 0.56 baris pertama merupakan nilai terhadap user scifi dan nilai -0.69, dan -0.69 merupakan nilai pada user romance.



'film', 'pengguna' dan 'konsep':

■ U: matriks kesamaan pengguna-ke-konsep

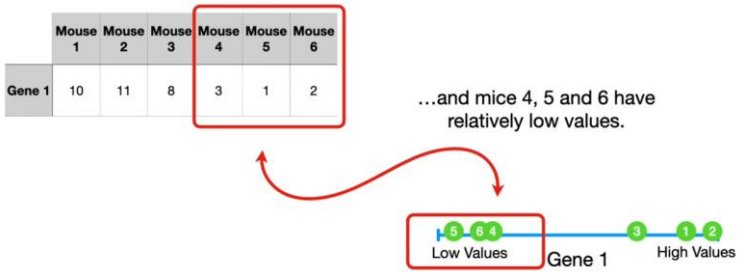
■ V: matriks kemiripan film dengan konsep

■ Σ: elemen diagonalnya:

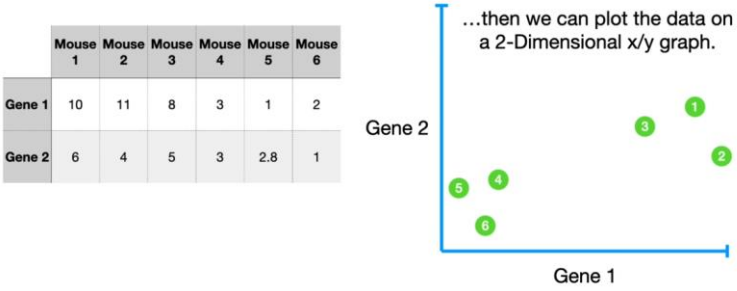
'kekuatan' setiap konsep

PCA

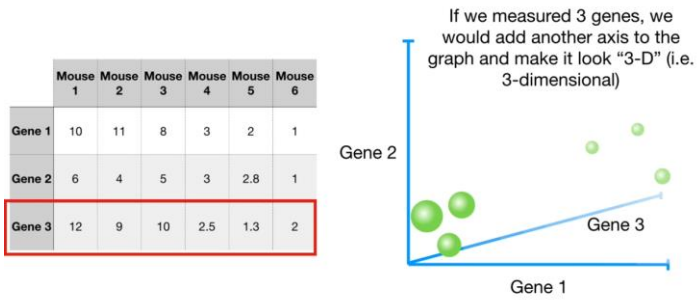
PCA (Principal Component Analysis) adalah sebuah teknik statistik yang digunakan untuk menganalisis dan mereduksi dimensi dari data multivariabel. Tujuan utama dari PCA adalah untuk mengidentifikasi pola dalam data dengan mengurangi jumlah variabel yang digunakan, sambil mempertahankan sebanyak mungkin informasi yang ada. PCA bekerja dengan mentransformasikan data asli ke dalam ruang dimensi yang lebih rendah yang disebut komponen utama (principal components). Komponen utama ini adalah linear kombinasi dari variabel asli, yang ditemukan sedemikian rupa sehingga komponen pertama (komponen utama pertama) memiliki varians yang paling tinggi, diikuti oleh komponen kedua dengan varians yang lebih rendah, dan seterusnya. Contohnya :



Jika table tersebut diukur dengan 1 gen, data dapat di plot dengan garis bilangan. Dengan isi kolom mouse yang berbeda mouse 1,2,3 memiliki nilai yang tinggi sedangkan sisanya rendah. Meskipun grafiknya sederhana, grafik ini menunjukkan kepada kita bahwa tikus 1, 2, dan 3 lebih mirip satu sama lain dibandingkan dengan tikus 4, 5, 6.

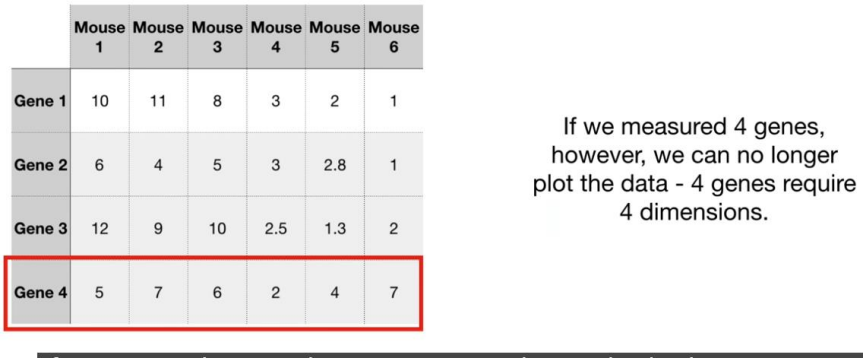


Pengukuran 2 gen dapat dilakukan dengan memplot ke 2 dimensi pada sumbu x dan y grafik. Pada isi tiap kolom table terdapat gen 1 dan 2, gen 1 merupakan nilai sumbu x dan gen 2 merupakan nilai sumbu y.



Pengukuran 3 gen dapat dilakukan dengan memplot ke 3 dimensi pada sumbu x (gen 1) dan y (gen 2), serta sumbu z (gen 2) grafik. Pada isi tiap kolom table terdapat gen 1, 2 dan 3, gen 1 merupakan nilai sumbu x dan gen 2 merupakan nilai sumbu y dan untuk gen 3 atau sumbu z adalah nilai yang

menentukan jauh dekat atau besar kecilnya suatu nilai misalnya pada mouse 1 memiliki nilai gen 1 = 10, gen 2 = 6, gen 3 = 12 dihasilkan titik atau bola dengan ukuran sedang sedangkan untuk mouse 6 memiliki nilai gen 1 = 1, gen 2 = 1, dan gen 3 = 2 dihasilkan titik atau bola yang besar dibanding sebelumnya.



	Mouse 1	Mouse 2	Mouse 3	Mouse 4	Mouse 5	Mouse 6
Gene 1	10	11	8	3	2	1
Gene 2	6	4	5	3	2.8	1
Gene 3	12	9	10	2.5	1.3	2
Gene 4	5	7	6	2	4	7

Pengukuran 4 gen dalam 4 dimensi tidak dapat dilakukan. Untuk memplot empat gen menggunakan PCA, pertama-tama, kita perlu mempersiapkan data yang mengandung ekspresi dari empat gen tersebut. Setiap gen akan menjadi dimensi dalam ruang empat dimensi. Data ekspresi gen ini perlu di-standarisasi jika skala ekspresi gen berbeda-beda. Setelah itu, PCA digunakan untuk mengidentifikasi komponen utama yang paling signifikan dalam data. Biasanya, hanya dua atau tiga komponen utama yang digunakan untuk tujuan visualisasi. Dalam plot PCA, data diproyeksikan ke dalam ruang yang lebih rendah, biasanya dua dimensi. Setiap titik dalam plot mewakili satu sampel atau observasi. Jika Anda memiliki informasi tambahan tentang sampel-sampel tersebut, Anda dapat memberikan warna atau simbol yang berbeda untuk membedakan kelompok atau kategori yang berbeda. Dengan melakukan ini, Anda dapat memvisualisasikan pola yang ada dalam data ekspresi gen keempat tersebut, dan mungkin mengidentifikasi keterkaitan antara sampel-sampel atau kelompok berdasarkan varian dalam data. PCA adalah alat yang kuat untuk mereduksi dimensi dalam analisis data genetik dan membantu Anda memahami hubungan antara empat gen dengan cara yang lebih terkelola dan intuitif.

LDA

LDA adalah singkatan dari "Linear Discriminant Analysis" (Analisis Diskriminan Linear) yang merupakan sebuah teknik statistik dan pembelajaran mesin yang digunakan dalam klasifikasi dan analisis data. Tujuan utama dari LDA adalah untuk memproyeksikan data dari ruang asli (biasanya dengan dimensi tinggi) ke dalam ruang yang lebih rendah sehingga memaksimalkan pemisahan antara kelas atau kelompok yang berbeda. LDA secara khusus dirancang untuk digunakan dalam tugas klasifikasi, di mana kita mencoba untuk mengklasifikasikan observasi atau sampel data ke dalam kategori atau kelas yang sesuai berdasarkan fitur-fiturnya.

Berbeda dengan PCA (Principal Component Analysis) yang bertujuan untuk mengurangi dimensi data dengan memaksimalkan varians total, LDA berfokus pada mengidentifikasi kombinasi linear dari variabel yang meminimalkan dispersi dalam setiap kelas (kelompok) dan memaksimalkan dispersi antara kelas. Hasil dari LDA adalah sejumlah komponen linier yang disebut "diskriminan" yang dapat digunakan untuk memisahkan kelas-kelas yang berbeda dalam data.

LDA sering digunakan dalam pemrosesan citra, pengenalan pola, dan tugas-tugas klasifikasi lainnya. Ini membantu dalam meningkatkan performa klasifikasi dengan mempertahankan informasi yang relevan sambil mengurangi dimensi data.

Dataset Iris mewakili 3 jenis bunga Iris (Setosa, Versicolour dan Virginica) dengan 4 atribut: panjang sepal, lebar sepal, panjang kelopak, dan lebar kelopak.

Analisis Komponen Utama (PCA) yang diterapkan pada data ini mengidentifikasi kombinasi atribut (komponen utama, atau arah dalam ruang fitur) yang paling banyak menyebabkan varians dalam data. Di sini kami memplot sampel yang berbeda pada 2 komponen utama pertama.

Analisis Diskriminan Linier (LDA) mencoba mengidentifikasi atribut yang paling banyak menimbulkan varian antar kelas. Secara khusus, LDA, berbeda dengan PCA, adalah metode yang diawasi, menggunakan label kelas yang dikenal.