

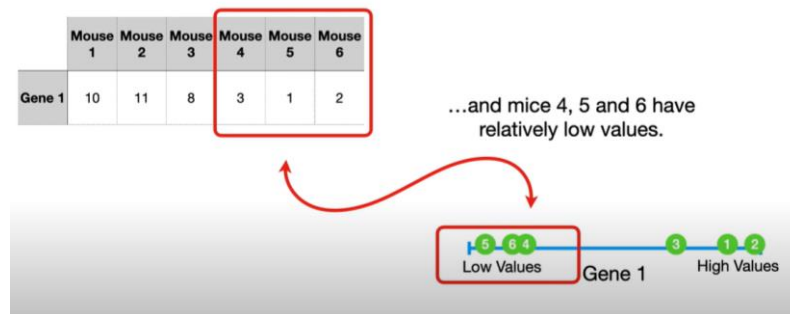
Analisis PCA

PCA adalah teknik statistika yang dapat digunakan untuk mengurangi dimensi data sambil tetap mempertahankan informasi yang signifikan. PCA bekerja dengan menemukan kombinasi linear dari variabel yang menjelaskan sebagian besar varians dalam data.

Contoh : Misalkan kita memiliki data tentang dua gen, Gen 1 dan Gen 2, pada enam tikus.

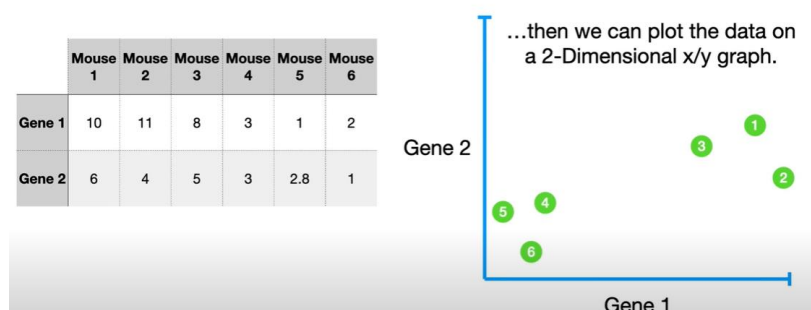
	Mouse 1	Mouse 2	Mouse 3	Mouse 4	Mouse 5	Mouse 6
Gene 1	10	11	8	3	1	2
Gene 2	6	4	5	3	2.8	1

Jika kita hanya mengukur satu gen, kita dapat memplot datanya pada garis bilangan.

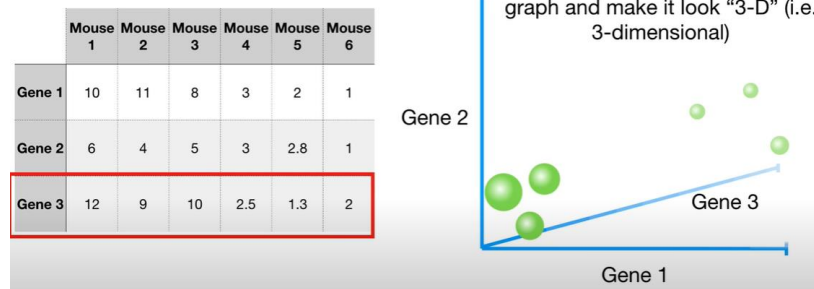


Jika kita mengukur dua gen, kita dapat memplot datanya pada grafik XY dua dimensi.

Gene 1 (X) dan Gen 2 (Y).



Jika kita mengukur tiga gen, kita dapat memplot datanya pada grafik 3D.



PCA dapat digunakan untuk mengurangi data ini menjadi dua dimensi. Untuk melakukan ini, PCA pertama-tama akan menemukan komponen utama pertama (PC1), yang adalah kombinasi linear dari dua gen yang menjelaskan sebagian besar varians dalam data. PC1 biasanya diproyeksikan ke sumbu x.

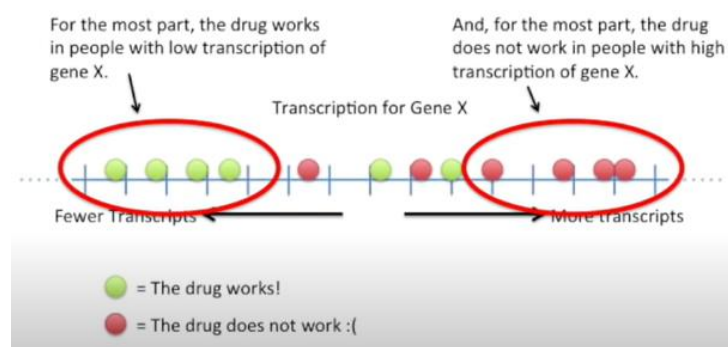
Nilai eigen: Nilai eigen adalah ukuran variasi yang dijelaskan oleh setiap komponen utama. Nilai eigen yang lebih besar menunjukkan bahwa komponen utama tersebut menjelaskan lebih banyak variasi.

Scree plot: Scree plot adalah representasi grafis dari persentase variasi yang dimiliki setiap komponen utama. Scree plot dapat digunakan untuk menentukan jumlah komponen utama yang perlu dipertahankan untuk mempertahankan informasi yang signifikan.

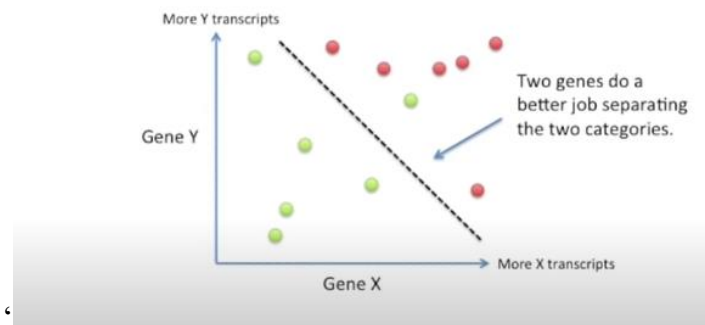
Analisis LDA

LDA adalah teknik statistika yang digunakan untuk memisahkan dua atau lebih kelompok data. LDA bekerja dengan membuat sumbu baru yang memaksimalkan jarak antara rata-rata kelompok sambil meminimalkan variasi dalam setiap kelompok. LDA dapat digunakan untuk berbagai tujuan, termasuk klasifikasi, prediksi, dan analisis tren.

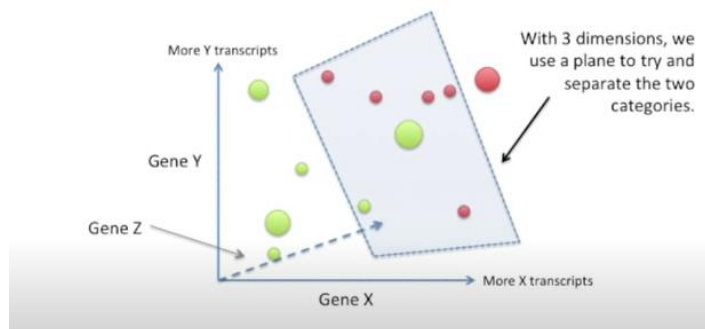
Using one gene to decide...



Using two genes to decide...



Using three genes to decide...



Dalam gambar yang terkait dengan video tersebut, terlihat dua kelompok data yang sangat relevan, yaitu kelompok berwarna hijau dan kelompok berwarna merah. Kelompok hijau menggambarkan pasien kanker yang telah sembuh dari penyakit, sementara kelompok berwarna merah mewakili pasien kanker yang belum pulih. Data ini terdiri dari dua fitur utama, yaitu tingkat transkripsi dari gen X dan gen Y.

Grafik pertama dalam dua dimensi mencerminkan bahwa kedua kelompok data tersebut memang terpisah, namun dengan beberapa area tumpang tindih. Tumpang tindih ini mengindikasikan adanya pasien kanker yang telah sembuh, tetapi memiliki tingkat transkripsi gen X yang mirip dengan pasien yang belum sembuh.

Grafik kedua dalam dua dimensi menunjukkan sebuah sumbu baru yang telah dibentuk oleh metode LDA (Linear Discriminant Analysis). Sumbu baru ini diciptakan dengan maksud untuk memaksimalkan jarak antara rata-rata tingkat transkripsi gen X dari kedua kelompok data.

Akibatnya, sumbu baru tersebut berhasil secara efektif memisahkan kedua kelompok data tersebut. Pasien kanker yang telah sembuh terletak di satu sisi sumbu, sementara pasien kanker yang belum pulih berada di sisi yang berlawanan.

Di dalam grafik tiga dimensi, terdapat sebuah titik data yang berwarna hitam yang mewakili pasien baru. Titik data ini terletak pada sisi yang sama dengan kelompok pasien kanker yang

telah sembuh. Dengan demikian, berdasarkan analisis LDA, kita dapat memprediksi bahwa pasien baru ini juga memiliki potensi untuk sembuh dari penyakit kanker.

Analisis SVD

Dalam video tersebut menjelaskan SVD (Singular Value Decomposition) adalah teknik untuk memecahkan matriks menjadi tiga matriks yang lebih kecil, yaitu U, Σ , dan V. Matriks U dan V mewakili arah-arrah utama dari data, sedangkan matriks Σ mewakili kekuatan dari arah-arrah tersebut.

$$A_{[m \times n]} = U_{[m \times r]} \Sigma_{[r \times r]} V_{[n \times r]}^T$$

Dengan penjelasan sebagai berikut:

Matriks A adalah matriks input data.

Matriks U adalah matriks yang mewakili vektor singular kiri.

Matriks Σ adalah matriks singular yang merupakan diagonal matrix dengan nilai-nilai singular yang mewakili kekuatan dari arah-arrah utama.

Matriks V adalah matriks yang mewakili vektor singular kanan.

SVD dapat digunakan untuk berbagai aplikasi, seperti:

Reduksi dimensi: SVD dapat digunakan untuk mengurangi dimensi data tanpa kehilangan informasi yang penting.

Pengenalan pola: SVD dapat digunakan untuk menemukan pola dalam data, seperti mengelompokkan data atau mengidentifikasi objek.

Pengolahan citra: SVD dapat digunakan untuk berbagai operasi pengolahan citra, seperti denoising, kompresi, dan segmentasi.

Pengolahan bahasa alami: SVD dapat digunakan untuk berbagai aplikasi pengolahan bahasa alami, seperti analisis sentimen, topik modeling, dan rekomendasi.