

NAMA : Syahrul Reza Ananda  
NIM : 1103210113

### StatQuest: Principal Component Analysis (PCA), Step-by-Step

PCA (Principal Component Analysis) adalah sebuah metode untuk mengurangi dimensi data dengan menggunakan Singular Value Decomposition (SVD). Dengan metode ini membantu mengatasi masalah ketika kita memiliki banyak dimensi dalam data, yang sulit untuk divisualisasikan. PCA memungkinkan kita untuk mengambil banyak pengukuran gen dan menghasilkan plot PCA dua dimensi. Ini membantu mengelompokkan data dan menyoroti gen mana yang paling berkontribusi terhadap perbedaan antar sampel.

Dalam langkah-langkah PCA, kita mencari garis-garis (komponen utama) yang paling baik memperjelas variasi dalam data. Dengan memutar sumbu untuk menghasilkan plot PCA akhir, kita dapat dengan mudah melihat bagaimana sampel-sampel terkelompok berdasarkan kesamaan atau perbedaan mereka. Penjelasan proses PCA dan cara menerapkannya dalam analisis data sebagai berikut:

- 1) Langkah pertama adalah menghitung rata-rata pengukuran untuk setiap gen dan menggeser data sehingga pusatnya berada di pusat grafik. Ini penting untuk memastikan bahwa data telah dikoreksi sehingga tidak terpengaruh oleh perbedaan dalam skala atau rata-rata antar variabel. PC1 mewakili arah dengan variabilitas terbesar dalam data. Misalnya, jika PC1 memiliki kemiringan 0,25, ini berarti data sebagian besar tersebar sepanjang sumbu Gen 1.
- 2) Komponen Utama 1 (PC1) adalah garis terbaik yang melalui asal setelah data digeser. Ini mewakili arah dengan variabilitas terbesar dalam data. Jika PC1 memiliki kemiringan 0,25, ini berarti data sebagian besar tersebar sepanjang sumbu Gen 1. Interpretasi ini membantu kita memahami sumbangan setiap variabel terhadap variasi keseluruhan dalam data.
- 3) Komponen Utama berikutnya (PC2) adalah garis terbaik yang tegak lurus dengan PC1. Ini membantu menangkap variasi tambahan dalam data yang tidak dijelaskan oleh PC1. Dengan memiliki PC2, kita dapat menangkap lebih banyak variasi dalam data yang tidak dapat dijelaskan oleh PC1 saja.
- 4) Plot PCA akhir dibuat dengan memutar PC1 menjadi sumbu horizontal dan PC2 menjadi sumbu vertikal. Titik-titik data kemudian diproyeksikan ke plot ini untuk menemukan lokasi sampel di dalamnya. Ini membantu kita memvisualisasikan sebaran data dalam ruang dua dimensi yang ditentukan oleh PC1 dan PC2.
- 5) PCA menggunakan Singular Value Decomposition (SVD) untuk menghitung PC dan mengurangi dimensi data. SVD adalah teknik matematika yang mendasari untuk menguraikan matriks data menjadi komponen utama yang merupakan kombinasi linier dari variabel-variabel asli.
- 6) Eigenvalues, yang dihitung dari hasil SVD, dapat digunakan untuk menentukan proporsi variasi yang dijelaskan oleh setiap PC. Scree plot, yang digambarkan dengan eigenvalues dari setiap PC, dapat digunakan untuk menganalisis variasi ini dan memutuskan berapa banyak komponen utama yang harus dipertahankan dalam analisis.