

Sam Alim Ramadhan
1103210236
TK-45-05

Principal Component Analysis (PCA) merupakan sebuah metode untuk mengurangi data dengan menggunakan Singular Value Decomposition (SVD). Metode ini membantu mengatasi masalah Ketika kita memiliki banyak dimensi dalam data yang sulit untuk divisualisasikan. Metode ini memungkinkan kita untuk mengambil banyak pengukuran gen dan menghasilkan plot PCA dua dimensi.

Dalam penerapannya, diawali dengan mencari garis-garis yang paling baik dalam memperjelas variasi pada data. Dengan memutar sumbu untuk menghasilkan plot PCA akhir, sampel-sampel terkelompok berdasarkan kesamaan atau perbedaan dapat dilihat dengan mudah. Berikut cara penerap PCA pada data analysis.

- kita bisa mulai dengan memplot data dan menghitung rata-rata pengukuran untuk setiap gen. Dengan nilai rata-rata tersebut, kita dapat menghitung pusat data. Setelah itu, kita fokus pada apa yang terjadi di grafik - kita tidak lagi membutuhkan data asli. Sekarang, kita geser data sehingga pusatnya berada di atas asal di grafik.
- Selanjutnya, kita mencoba untuk mencocokkan garis pada data tersebut. Untuk melakukan ini, kita mulai dengan menggambar garis acak yang melewati asal, lalu kita memutar garis tersebut hingga cocok dengan data sebaik mungkin. Akhirnya, kita dapat mengukur jarak dari titik data ke garis tersebut dan mencari garis yang memiliki jarak terbesar. Garis ini disebut Komponen Utama 1, atau PC1. PC1 memiliki kemiringan 0,25, yang berarti data sebagian besar tersebar sepanjang sumbu Gen 1, dan hanya sedikit tersebar sepanjang sumbu Gen 2. Satu cara untuk memahami PC1 adalah dalam bentuk resep koktail. Untuk membuat PC1, campur empat bagian Gen 1 dengan satu bagian Gen 2.
- Selanjutnya, kita mencari PC2, yang merupakan garis terbaik berikutnya yang tegak lurus dengan PC1 tanpa optimasi lebih lanjut yang harus dilakukan. PC2 adalah -1 bagian Gen 1 ke 4 bagian Gen 2. Akhirnya, kita bisa menggambar plot PCA akhir dengan memutar sehingga PC1 horizontal dan PC2 vertikal, dan kemudian menggunakan titik-titik yang diproyeksikan untuk menemukan lokasi sampel di plot PCA.
- Demikianlah cara PCA dilakukan menggunakan Singular Value Decomposition. Selanjutnya, kita bisa menggunakan eigenvalues untuk menentukan proporsi variasi yang diperoleh setiap PC, dan membuat scree plot untuk menganalisis variasi tersebut.