

Nama : Jhose Immanuel Sembiring
NIM : 1103202047
Kelas : TK 2324-2

PCA

Principal Component Analysis (PCA) adalah sebuah teknik statistik yang digunakan untuk mengurangi dimensi dari dataset yang kompleks, sambil mempertahankan informasi yang paling penting dalam data. PCA bekerja dengan mentransformasi data ke dalam sistem koordinat baru yang disebut sebagai ruang komponen utama (principal component space), di mana varians data tersebar maksimal di sepanjang komponen utama yang pertama, diikuti oleh komponen utama kedua, dan seterusnya.

Tujuan utama dari PCA:

- Mengidentifikasi pola tersembunyi dalam data dengan mengurangi dimensi dari dataset, sehingga memungkinkan pemahaman yang lebih baik terhadap struktur data dan memfasilitasi analisis yang lebih efektif.
- Membantu dalam mengurangi overfitting dalam model statistik, mempermudah visualisasi data yang kompleks, serta memungkinkan pemahaman yang lebih baik terhadap hubungan antar variabel dalam dataset.

PCA sering digunakan dalam berbagai bidang seperti analisis data, pengenalan pola, pembelajaran mesin, analisis citra, dan berbagai aplikasi lainnya di mana pengurangan dimensi data yang kompleks diperlukan untuk mempermudah interpretasi dan analisis.

SVD(Singular Value Decomposition)

Merupakan konsep dalam aljabar linear yang digunakan untuk menganalisis dan merepresentasikan matriks.

PCA sering kali diimplementasikan menggunakan SVD. Dalam konteks PCA, SVD digunakan untuk menghitung komponen-komponen utama dari dataset. Dalam proses ini, SVD digunakan untuk mendekomposisi matriks kovarian dari dataset menjadi matriks-matriks yang merepresentasikan arah-arah utama variasi dalam data.

Langkah-langkah Visualisasi Data

1. Data yang akan divisualisasikan harus disiapkan terlebih dahulu.
2. Memahami dimensi data yang akan divisualisasikan.
3. Menggunakan Principal Component Analysis (PCA) untuk mengurangi dimensi data. PCA membantu dalam mengidentifikasi hubungan antara variabel dan memperjelas struktur data.
4. Menggunakan SVD dalam PCA untuk memisahkan variabel-variabel yang saling terkait.
5. Memplot data ke dalam grafik PCA. Ini dapat dilakukan dengan menggunakan dua komponen utama yang dihasilkan dari PCA.
6. Mengidentifikasi komponen utama (Principal Components) dari data. PC1 adalah garis terbaik yang dapat disesuaikan dengan data yang memiliki variasi terbesar, diikuti oleh PC2, PC3, dan seterusnya.
7. Menginterpretasikan hasil visualisasi PCA untuk mendapatkan pemahaman yang lebih dalam tentang data.

8. Membuat scree plot untuk melihat seberapa banyak variasi yang dijelaskan oleh setiap komponen utama.
9. Berdasarkan scree plot, memilih jumlah dimensi yang akan dipertahankan untuk visualisasi.
10. Menggunakan komponen utama yang dipilih untuk memproyeksikan data ke dalam dimensi yang lebih rendah.
12. Menampilkan data yang telah diproyeksikan ke dalam grafik PCA yang baru.
13. Menginterpretasikan grafik PCA untuk mendapatkan pemahaman yang lebih baik tentang pola atau hubungan dalam data.