

NAMA : Alfin Andreas Bastian Situmeang
NIM : 1103202143

PCA adalah singkatan dari Principal Component Analysis. Ini adalah metode statistik yang digunakan untuk mereduksi dimensi dari dataset yang kompleks, sementara mempertahankan sebagian besar informasi yang ada. PCA adalah alat yang sangat berguna dalam analisis data, terutama ketika bekerja dengan dataset yang memiliki banyak fitur atau variable.

Secara intuitif, PCA berusaha untuk menemukan pola atau struktur yang ada dalam data dengan menemukan arah terpenting atau komponen utama dari variasi dalam data. Ini dilakukan dengan mentransformasi data ke ruang dimensi yang lebih rendah, di mana komponen-komponen utama ini menjadi sumbu utama baru. Proses ini dilakukan sedemikian rupa sehingga komponen-komponen baru ini secara berurutan memaksimalkan variasi data yang tersisa. Dengan cara ini, PCA memungkinkan kita untuk mereduksi dimensi dari data tanpa kehilangan terlalu banyak informasi.

Langkah-langkah utama dalam PCA adalah sebagai berikut:

Standarisasi data: Pertama, data yang akan dianalisis harus disesuaikan sehingga setiap variabel memiliki mean 0 dan varians 1. Ini penting karena PCA sensitif terhadap skala variabel.

Perhitungan matriks kovariansi: Setelah data telah distandarisasi, matriks kovariansi dari variabel-variabel tersebut dihitung. Matriks kovariansi memberikan informasi tentang seberapa berhubungan variabel-variabel dalam dataset.

Perhitungan komponen-komponen utama: Komponen-komponen utama dari matriks kovariansi dihitung. Ini adalah vektor yang menunjukkan arah terpenting dari variasi dalam data.

Proyeksi data ke ruang dimensi yang lebih rendah: Data kemudian diproyeksikan ke ruang yang terdiri dari komponen-komponen utama yang paling signifikan, yang disebut sebagai ruang fitur baru.

Interpretasi hasil: Setelah proses PCA selesai, hasilnya dapat diinterpretasikan untuk memahami struktur dalam data. Komponen utama dengan nilai eigen yang tinggi menyimpan lebih banyak informasi tentang variasi dalam data.

PCA memiliki banyak aplikasi dalam berbagai bidang, termasuk pengenalan pola, pengolahan citra, pengenalan wajah, bioinformatika, dan lain-lain. Ini adalah alat yang sangat berguna untuk mengatasi masalah dimensi dalam analisis data.