

Nama : Muhammad Raihan Butar-Butar

NIM : 1103213077

Principal Component Analysis

Principal Component Analysis atau PCA adalah sebuah teknik statistik yang digunakan untuk mereduksi dimensi dari dataset yang kompleks dengan tujuan untuk memperjelas struktur data serta mempertahankan sebanyak mungkin informasi yang dapat dipertahankan. PCA mencari kombinasi linear dari variabel asli yang menjelaskan variasi maksimum dalam data.

Proses PCA dimulai dengan menghitung matriks kovariansi atau korelasi dari dataset. Kemudian PCA mengidentifikasi vektor-vektor eigen dari matriks kovariansi ini yang merepresentasikan arah dari variabilitas maksimum dalam data. Komponen-komponen utama ini disusun berdasarkan nilai eigen yang terkait, di mana komponen pertama memiliki nilai eigen yang paling besar, diikuti oleh komponen kedua, dan seterusnya.

Setelah itu PCA mentransformasikan data ke dalam ruang baru yang terdiri dari komponen-komponen utama. Transformasi ini memungkinkan kita untuk merepresentasikan data dengan dimensi yang lebih rendah, dengan mempertahankan sebagian besar informasi dari dataset asli.

Selama transformasi, komponen-komponen yang memiliki kontribusi terbesar terhadap variasi data dipertahankan, sementara yang memiliki kontribusi yang lebih kecil dibuang. Oleh karena itu, PCA membantu dalam menghilangkan redundansi dan mengurangi noise dalam dataset.

Terakhir, hasil dari PCA dapat digunakan untuk menganalisis hubungan antar variabel, menemukan pola dalam data, serta memvisualisasikan data dengan dimensi yang lebih rendah, yang memungkinkan interpretasi yang lebih mudah dan pemahaman yang lebih baik terhadap struktur data. Dengan demikian, PCA merupakan alat yang kuat dalam analisis data untuk mengeksplorasi dan memahami struktur intrinsik dalam dataset yang kompleks.