

Nama : Taufik Kurahman Sasmita

Kelas : TK 2325-2

Nim : 1103200008

PCA (Principal Component Analysis)

PCA (Principal Component Analysis) adalah teknik dalam machine learning yang berguna untuk mengurangi dimensi dari dataset yang kompleks. Tujuan utamanya adalah untuk mengidentifikasi pola dalam data dengan cara mengurangi jumlah variabel yang ada, sambil mempertahankan sebanyak mungkin informasi yang relevan. PCA menciptakan kombinasi linear baru dari variabel asli, yang disebut sebagai komponen utama, sehingga menghasilkan ruang dimensi yang lebih kecil.

Proses PCA dimulai dengan menghitung matriks kovariansi dari dataset untuk menemukan hubungan antar variabel. Selanjutnya, dilakukan dekomposisi nilai eigen dari matriks kovariansi ini untuk mendapatkan vektor eigen dan nilai eigen. Vektor eigen ini menunjukkan arah di mana data tersebar dengan maksimal, sedangkan nilai eigen mengukur seberapa banyak variabilitas yang dijelaskan oleh setiap vektor eigen.

Setelah mendapatkan nilai eigen, PCA memilih komponen utama berdasarkan nilai eigen terbesar, karena komponen-komponen ini mengandung sebagian besar variabilitas dalam data. Dengan memilih hanya sejumlah komponen utama tertentu, dimensi dataset dapat dikurangi tanpa kehilangan terlalu banyak informasi.

Dengan menerapkan PCA, kita dapat mengurangi kompleksitas data, mempercepat proses pembelajaran model, dan memperbaiki kinerja model machine learning dengan menghilangkan varians yang tidak relevan atau redundan. Ini membuat PCA menjadi alat yang sangat berguna dalam analisis data dan pengembangan model machine learning.