

Nama : Mochamad Phillia Wibowo

NIM : 1103204191

Kelas : Machine Learning (TK-44-G04)

Tugas Lecture 5 : Membuat Catatan PCA

Principal Component Analysis (PCA) adalah teknik statistik yang membantu kita memahami data dengan lebih dalam. PCA bekerja dengan menyederhanakan data ke dalam ruang yang lebih kecil, sehingga kita dapat dengan mudah melihat pola dan perbedaan di antara elemen-elemen datanya. Misalnya, dalam sebuah penelitian yang mengamati ekspresi dua gen pada enam tikus yang berbeda, PCA membantu kita mengenali pola-pola yang mungkin ada di antara tikus-tikus tersebut.

Pertama-tama, PCA mengubah data ke dalam grafik satu atau dua dimensi. Misalnya, pada grafik satu dimensi, kita dapat melihat perbedaan antara tikus-tikus berdasarkan nilai-nilai genetik mereka. Selanjutnya, PCA menggunakan rata-rata data untuk menemukan pusat dari distribusi data, dan kemudian memproyeksikan data ke dalam garis yang paling sesuai dengan pola-pola tersebut.

Selama proses ini, PCA memperhatikan jarak antara titik-titik data dengan garis proyeksi atau jarak antara titik proyeksi dengan pusat grafik. PCA berusaha mencari garis yang paling sesuai dengan data dengan memaksimalkan jarak antara titik proyeksi dan pusat grafik. Garis ini disebut sebagai Komponen Utama 1 (PC1) dan membantu kita melihat sebaran data dalam dua dimensi, misalnya di sepanjang sumbu Gen 1 dan Gen 2.

Selanjutnya, PCA melakukan penyesuaian skala agar garis merah yang mewakili PC1 memiliki panjang satu unit. Ini membantu kita mengidentifikasi proporsi masing-masing gen dalam vektor singular atau vektor eigen untuk PC1. PC2, yang tegak lurus dengan PC1, juga diidentifikasi dalam proses ini.

Eigenvalues digunakan sebagai ukuran variasi dalam PCA, yang memberi tahu kita seberapa banyak informasi yang dijelaskan oleh setiap komponen utama. PC1 dan PC2, misalnya, masing-masing mewakili sebagian besar variasi dalam data. Grafik scree membantu kita memvisualisasikan kontribusi masing-masing komponen utama terhadap variasi data.

Dengan demikian, PCA menyimpulkan bahwa PC1 dan PC2 sudah cukup untuk mewakili sebagian besar variasi dalam data. Ini memungkinkan kita membuat grafik PCA 2D yang sederhana namun informatif untuk memahami pola data dengan lebih baik.