

Nama : Mochamad Permana Ash Shidiq
NIM : 1103228233

PCA StatQuest

PCA atau Principal Component Analysis adalah teknik analisis data yang digunakan untuk mengurangi kompleksitas kumpulan data dengan mengurangi jumlah variabel menjadi beberapa faktor utama. Hal ini membantu dalam memvisualisasikan data dan menentukan variabel mana yang paling penting untuk memahami kumpulan data secara keseluruhan.

Proses PCA dimulai dengan memproyeksikan titik-titik data ke dalam sebuah garis, kemudian mengukur jarak antara titik-titik tersebut dan garis tersebut. Garis dengan jumlah kuadrat jarak terbesar antara titik proyeksi dan titik asal disebut komponen utama (PC1). Dengan mengurutkan komponen utama berdasarkan jumlah varians yang dijelaskan, PCA memungkinkan kita untuk memahami bagaimana variabilitas dalam data didistribusikan.

PCA juga dapat digunakan untuk mereduksi jumlah dimensi dalam kumpulan data. Ini dilakukan dengan menggabungkan berbagai bagian data (misalnya, sampel gen) dan kemudian menghitung kombinasi linier dari sampel tersebut yang menghasilkan nilai tunggal terbesar. Nilai tunggal ini, disebut nilai eigen, digunakan untuk menghitung vektor eigen untuk plot PCA, yang membantu menentukan mana yang paling penting dalam mendeskripsikan data.

StatQuest memberikan penjelasan yang jelas tentang bagaimana menghitung komponen utama dan bagaimana menggunakan informasi ini untuk membuat plot data 2D yang disederhanakan. Mereka juga memberikan tutorial tentang cara mengubah plot 2D menjadi grafik 3D agar lebih mudah dibaca.

Dengan menggunakan PCA, kita dapat memahami struktur dalam data dengan cara yang lebih sederhana dan mudah dipahami. Hal ini memungkinkan kita untuk mengambil keputusan yang lebih baik berdasarkan pemahaman yang lebih baik tentang variabilitas dan hubungan dalam data.