

PCA

PCA Adalah sebuah teknik untuk menyederhanakan dimensi dari sebuah dataset yang kompleks, sambil mempertahankan informasi penting yang ada di dalam dataset tersebut. PCA bertujuan untuk menemukan insight dan pola yang tersembunyi di dalam dataset dengan cara mentransformasi variabel asli menjadi variabel yang tidak berkorelasi, yang disebut sebagai komponen utama, komponen utama ini merupakan kombinasi linear dari variabel asli.

secara sederhana, PCA terdiri dari langkah langkah :

1. menggeser data agar rata rata berada di tengah tengah grafik
2. mencari garis yang terbaik dalam merepresentasikan data dengan cara mengkalkulasikan SS(sum of squared distance), dan memilih garis dengan SS terbesar, garis ini dinamakan PC(Principal Component).
3. tentukan data mana yang lebih penting dari informasi yang didapat saat mencari PC1
4. garis PC2 dicari dengan cara menarik garis dari titik tengah grafik dan harus tegak lurus dari garis PC1.
5. lakukan rotasi terhadap garis PC1 dan PC2 sehingga masing masing garis menjadi sumbu x dan sumbu y.
6. kalkulasikan projected point dari tiap sampel untuk mencari letak sampel di grafik yang baru, projected point ini didapat dengan cara menghitung eigenvalue dan garis PC1.

PCA banyak digunakan dalam berbagai bidang seperti analisis data, machine learning, pengolahan citra, dan lainnya untuk mengatasi masalah dimensi tinggi dan mengekstraksi fitur-fitur yang paling penting dari dataset.