

Nama : Angelica Sharon Amelia Simanjuntak
NIM : 1103210032
Tugas Minggu 5 : Membuat catatan PCA

StatQuest: Principal Component Analysis (PCA), Step-by-Step

Di dalam video ini membahas Principal Component Analysis (PCA), selangkah demi selangkah dengan menggunakan Singular Value Decomposition (SVD). Principal Component Analysis (PCA) adalah metode reduksi dimensi dan machine learning yang digunakan untuk menyederhanakan kumpulan data besar menjadi kumpulan yang lebih kecil sambil tetap mempertahankan pola dan tren yang signifikan. Di dalam video dijelaskan PCA berfungsi untuk membantu dalam transkripsi dua gen terhadap enam tikus yang berbeda. Gen tikus yang berbeda ini dijadikan sebagai representasi variabel yang nantinya akan diukur pada setiap tikus yang menjadi sampel. Di dalam contoh ini, PCA membantu kita dalam memahami Gen mana yang paling penting saat mengelompokkan data. Pada video ini diketahui bahwa Gen 3 yang memiliki peran penting dalam memisahkan sampel karena gen 3 letaknya lebih jauh dan titik yang lebih kecil memiliki nilai yang lebih besar.

Setelah kita mengetahui komponen utama, kita menggunakan nilai eigen, yaitu jumlah kuadrat jarak, untuk menentukan proporsi variasi yang diperhitungkan setiap PC. Dalam hal ini, PC1 menyumbang 79% variasi, PC2 menyumbang 15% variasi, dan PC3 menyumbang 6% variasi. Berikut plot layarnya. PC1 dan PC2 menyumbang sebagian besar variasi. Artinya, grafik 2D, yang hanya menggunakan PC1 dan PC2, merupakan perkiraan yang baik untuk grafik 3D ini, karena grafik tersebut mencakup 94% variasi data. Untuk mengubah grafik 3D menjadi grafik PCA dua dimensi, kita cukup menghapus semuanya kecuali data dan PC1 dan PC2, lalu memproyeksikan sampelnya ke PC1 dan PC2. Lalu kita putar sehingga PC1 horizontal dan PC2 vertikal. Ini hanya membuatnya lebih mudah untuk dilihat. Karena titik-titik yang diproyeksikan ini sesuai dengan sampel 4, di sinilah sampel 4 masuk dalam plot PCA baru.

Menggambarkan grafik PCA dapat dilakukan dengan langkah-langkah berikut, langkah yang pertama menghitung nilai komponen utama, lalu memilih komponen utama, yang ketiga membuat scatter plot, dan yang keempat melakukan interpretasi.

