

Nama : M Rayyan Aqiilah M

NIM : 1103210210

Kelas : TK-44-G4

Rangkuman StatQuest PCA

Video dimulai dengan memberi penjelasan menggunakan dataset sederhana. Jika kita melakukan plotting terhadap 1 jenis data, maka hasil plotting yang muncul hanyalah 1 garis. Jika kita melakukan plotting terhadap 2 dan 3 jenis data maka visualisasi yang muncul adalah seperti scatter plot. Tetapi kita tidak akan dapat melakukan plotting terhadap 4 jenis data atau lebih. Oleh karena itu, kita akan menggunakan PCA.

PCA dimulai dengan menghitung average measurement masing masing jenis data. Dengan data yang didapat kita dapat menghitung center of data. Center of data akan ditempatkan di tengah lalu kita akan menggambar garis lalu memutarinya sampai garis tersebut tepat untuk persebaran data yang kita miliki.

Untuk mencari PC1, kita akan menghitung jarak persebaran masing masing data kepada garis imajiner yang kita gambar dengan catatan dikuadratkan agar nilai negatif tidak mengganggu proses perhitungan. Setelah itu kita akan menjumlahkan seluruh angka yang kita dapatkan untuk mendapat nilai SS. Lalu kita ulangi menggunakan garis imajiner lain sampai menemukan nilai SS terbesar dimana nilai tersebut akan dijadikan PC1. Untuk mendapatkan nilai Eigenvalue PC1, kita dapat membagi SS dengan $n-1$, sedangkan untuk singular value PC1, kita dapat meng akarkan nilai SS PC1.

Langkah selanjutnya adalah mencari PC2. Kita dapat mendaptkan PC2 dengan cara membuat garis yang tegak lurus terhadap garis PC1, lalu garis tersebut adalah PC2. Setelah mendapatkan PC1 dan PC2, kita akan menjadikan PC1 dan PC2 menjadi grafik baru dengan PC1 sebagai sumbu x dan PC2 sebagai sumbu y. Untuk mencari variasi persen dari PC1 dan PC2 kita dapat membandingkan eigenvalue dari keduanya.