

Nama : Muhammad Ashlah Afifi

NIM : 1103210216

Kelas : TK-44-G4

## StatQuest PCA

Pada awal Video dijelaskan bahwa Ketika A, dan A dikuadratkan, maka ketika dan B bertambah besar maka C harus mengecil. Begitu pula jika C semakin besar maka B harus semakin kecil. Dengan demikian, PCA dapat meminimalkan jarak ke garis, atau memaksimalkan jarak dari titik proyeksi ke titik asal. Alasan saya mempermasalahkan hal ini adalah, secara intuitif, masuk akal untuk meminimalkan B dan jarak dari titik ke garis, namun sebenarnya lebih mudah untuk menghitung C, jarak dari titik yang diproyeksikan ke titik asal, jadi PCA menemukan garis yang paling cocok dengan memaksimalkan jumlah kuadrat jarak dari titik yang diproyeksikan ke titik asal. Jadi, untuk garis ini, PCA memproyeksikan data ke dalamnya dan kemudian mengukur jarak dari titik ini ke titik asal, sebut saja D1. Dan kemudian PCA mengukur jarak dari titik ini ke titik asal, kita menyebutnya D2. Kemudian diukur D3, D4, D5, dan D6. Berikut enam jarak yang kami ukur.

Hal berikutnya yang kita lakukan adalah mengkuadratkan semuanya. Jaraknya dikuadratkan sehingga nilai negatif tidak menghilangkan nilai positif. Lalu kita jumlahkan semua jarak kuadrat tersebut, dan hasilnya sama dengan jumlah jarak kuadrat. Singkatnya kita akan menyebutnya jarak SS, untuk jumlah jarak kuadrat. Sekarang kita memutar garis, memproyeksikan data ke garis, dan kemudian menjumlahkan jarak kuadrat dari titik yang diproyeksikan ke titik asal. Dan kita ulangi sampai kita mendapatkan garis dengan jumlah kuadrat jarak terbesar antara titik yang diproyeksikan dan titik asal.

Pada akhirnya, kita berakhir dengan garis ini. Ia mempunyai jumlah kuadrat jarak terbesar. Baris ini disebut Komponen Utama 1, atau disingkat PC1. PC1 memiliki kemiringan 0,25. Dengan kata lain, untuk setiap 4 unit yang kita keluarkan di sepanjang sumbu Gen 1, kita naik 1 unit di sepanjang sumbu Gen 2. Artinya sebagian besar data tersebar di sepanjang sumbu Gen 1, dan hanya sedikit tersebar di sepanjang sumbu Gen 2. Salah satu cara untuk memikirkan PC1 adalah dari segi resep koktail.

Untuk membuat PC1, campurkan empat bagian Gene 1 dengan satu bagian Gene 2. Tuangkan di atas es dan sajikan! Rasio Gen 1 dan Gen 2 memberi tahu Anda bahwa Gen 1 lebih penting dalam mendeskripsikan penyebaran data. Oh tidak, peringatan terminologi! Para ahli matematika menyebut resep koktail ini sebagai kombinasi linier dari Gen 1 dan 2. Saya menyebutkan ini karena ketika seseorang mengatakan PC1 adalah kombinasi variabel linier, inilah yang mereka bicarakan. Ini bukan masalah besar. Resep untuk PC1, melampaui 4 dan naik 1 membawa kita ke titik ini. Kita bisa mencari panjang garis merah dengan menggunakan teorema Pythagoras, A kuadrat yang lama sama dengan B kuadrat ditambah C kuadrat.

Memasukkan angka-angka tersebut memberi kita A sama dengan 4,12. Jadi panjang garis merahnya adalah 4,12. Ketika melakukan PCA dengan SVD, resep untuk PC1 akan diskalakan sehingga panjangnya sama dengan 1. Yang harus kita lakukan untuk menskalakan segitiga sehingga garis merah menjadi 1 satuan panjang yaitu dengan membagi setiap sisinya dengan 4,12. Kita masih menggunakan Gen 1 empat kali lebih banyak dibandingkan Gen 2. Jadi

sekarang kita kembali melihat data, garis yang paling cocok, dan vektor satuan yang baru saja kita hitung.