

Nama : Muhammad Alfi Syahrin
NIM : 1103213144

Catatan PCA

Pada suatu data, kita dapat memiliki beberapa variabel yang mengidentifikasi sampel-sampel. Jika terdapat satu, dua, atau tiga variabel kita dapat membuat grafiknya. Namun, jika variabel lebih dari tiga, kita sudah tidak dapat melakukan visualisasinya. Pada video ini, kita dapat mengetahui bagaimana cara PCA melakukan perhitungan dan membuat visualisasi data dalam bentuk dua dimensi (PCA Plot). Melalui video ini juga kita mempelajari bagaimana PCA dapat memberitahui variabel-variabel yang paling penting untuk melakukan pengelompokan data.

Cara PCA Bekerja pada 2 Variabel

Dimulai dengan melakukan plotting data, setelah itu dilakukan perhitungan rata-rata untuk tiap variabel. Melalui rata-rata tersebut, kita melakukan perhitungan untuk mencari titik tengah dari data. Setelah itu dilakukan pergeseran data dengan titik tengah menjadi titik (0,0). Setelah itu, kita dapat membuat garis yang sesuai untuk data tersebut dengan cara membuat garis acak dari garis asal dan melakukan perputaran sehingga garis paling sesuai dengan data tersebut.

Bagaimana PCA Menentukan Garis Paling Sesuai

Setelah menemukan garis yang sesuai dengan data, PCA mengukur jarak dari setiap titik data ke garis tersebut. Kemudian, PCA dapat memutuskan apakah garis tersebut baik atau tidak dengan dua metode: meminimalkan jarak antara titik data dan garis, atau memaksimalkan jarak antara titik proyeksi dan titik asal.

Dalam prakteknya, PCA memilih metode yang memaksimalkan jarak antara titik proyeksi dan titik asal, karena lebih mudah dihitung. Setelah menentukan garis terbaik (Komponen Utama 1 atau PC1), PCA menghitung nilai eigen dan vektor singular untuk garis tersebut.

Nilai eigen mewakili seberapa banyak variasi yang dijelaskan oleh PC1, sedangkan vektor singular (atau loading scores) menunjukkan kontribusi masing-masing variabel terhadap PC1. Proses ini dilanjutkan untuk menemukan Komponen Utama kedua (PC2), yang tegak lurus dengan PC1 dan juga melewati titik asal. Dari hal tersebut, PCA menghasilkan dua Komponen Utama yang menjelaskan sebagian besar variasi data dan dilakukan visualisasi data berdasarkan PC1 dan PC2.