

PCA(Principal Component Analysis)

1. Cara kerja Algoritma

Secara garis besar, algoritma PCA terdiri dari tiga tahap utama: mean normalization, covariance matrix computation, dan eigen decomposition.

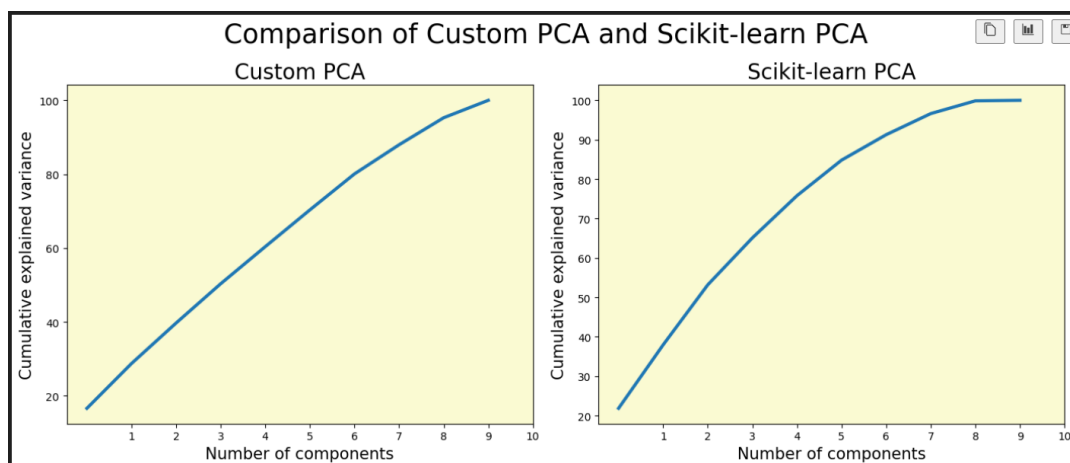
Pada tahap pertama, setiap fitur dalam dataset akan dinormalisasi dengan cara mengurangi nilai rata-rata (mean) dari setiap nilai fitur tersebut. Ini bertujuan untuk memastikan bahwa data berpusat di sekitar asal (0) sehingga analisis lebih akurat.

Pada tahap kedua, matriks kovarians akan dihitung berdasarkan data yang sudah dinormalisasi. Matriks kovarians ini menggambarkan bagaimana dua variabel berhubungan satu sama lain di dalam dataset.

Pada tahap ketiga, dilakukan dekomposisi eigen terhadap matriks kovarians untuk mendapatkan eigenvectors (komponen utama) dan eigenvalues (nilai eigen). Eigenvectors mewakili arah dari varians terbesar dalam data, sedangkan eigenvalues menunjukkan seberapa besar varians tersebut.

Setelah mendapatkan eigenvectors, kita dapat memproyeksikan data asli ke dalam ruang baru yang terbentuk oleh komponen utama tersebut. Umumnya, hanya beberapa komponen utama pertama yang digunakan, tergantung pada jumlah varians yang ingin dipertahankan. Algoritma ini diulang-ulang atau diaplikasikan hingga jumlah komponen utama yang diinginkan tercapai.

2. Perbandingan Hasil Evaluasi Model



Hasil explained variance memiliki perbedaan yang cukup banyak. Hal ini terjadi karena model dari scikit learn menggunakan Singular Value Decomposition yang lebih stabil daripada Eigen Decomposition yang digunakan oleh model saya. Selain itu, model saya melakukan scaling pada saat fitting sedangkan model dari scikit learn tidak melakukan hal itu.