

Nama : Rizka Rahmadina

NIM : 1103204115

## Understanding PCA, LDA, SVD and Pandas Framework

- PCA

Principal Component Analysis (PCA) adalah sebuah metode statistik yang digunakan untuk mengurangi dimensi data dalam analisis multivariat. PCA bertujuan untuk mengidentifikasi dan menjelaskan variasi dalam data dengan mengubah variabel asli ke ruang dimensi yang lebih rendah, yang disebut sebagai komponen utama. Langkah pertama dalam PCA adalah menghitung matriks kovariansi dari data asli, yang mengukur hubungan antara variabel. Kemudian, PCA menghitung vektor dan nilai eigen dari matriks kovariansi ini. Komponen utama adalah vektor eigen yang menjelaskan variasi terbesar dalam data. Dengan mengurutkan komponen utama berdasarkan nilai eigen, kita dapat memilih sejumlah komponen utama yang akan digunakan untuk menggantikan data asli dalam dimensi yang lebih rendah.

- LDA

Linear Discriminant Analysis (LDA), di sisi lain, adalah teknik yang digunakan dalam analisis statistik dan pembelajaran mesin untuk meningkatkan pemisahan kelas dalam data. LDA adalah metode pengurangan dimensi yang berfokus pada pemisahan kelas, berbeda dengan PCA yang hanya berfokus pada variasi total. LDA bekerja dengan menghitung matriks scatter antar kelas dan matriks scatter dalam kelas. Tujuannya adalah untuk mengoptimalkan pemisahan antar kelas sedangkan meminimalkan pemisahan dalam kelas. Hasilnya adalah vektor diskriminan atau komponen diskriminan yang digunakan untuk mentransformasi data ke dimensi yang lebih rendah, dengan mempertahankan sebanyak mungkin informasi kelas.

- SVD

Singular Value Decomposition (SVD) adalah teknik dekomposisi matriks yang digunakan dalam berbagai bidang, termasuk matematika, statistik, dan ilmu data. SVD mengambil matriks dan membaginya menjadi tiga matriks: matriks singular value, matriks left singular, dan matriks right singular. SVD digunakan dalam PCA untuk menghitung komponen utama dan dalam berbagai aplikasi lain seperti reduksi dimensi, kompresi data, dan pemrosesan sinyal. Dalam konteks PCA, SVD digunakan untuk mendekomposisi matriks kovariansi data, dan komponen utama dihasilkan dari matriks left singular. SVD juga berguna dalam analisis matriks termasuk rekonstruksi data asli dari data yang telah direduksi dimensi.

- Pandas Framework

Pandas adalah sebuah framework open-source yang populer untuk analisis data dalam bahasa pemrograman Python. Framework ini menyediakan alat yang kuat untuk mengelola dan memanipulasi data tabular, seperti spreadsheet atau basis data SQL, sehingga sangat berguna dalam analisis data, ilmu data, dan ilmu komputer secara umum. Pandas mengenalkan dua struktur data utama: Series dan DataFrame. Series adalah struktur data satu dimensi yang mirip dengan array atau list, sementara DataFrame adalah struktur data dua dimensi yang menyerupai tabel dengan baris dan kolom.

Pandas memberikan kemampuan untuk mengimpor data dari berbagai sumber, seperti file CSV, Excel, SQL databases, dan bahkan web, sehingga memudahkan pengguna dalam memuat data ke dalam lingkungan

Python. Setelah data dimuat, Pandas memungkinkan pengguna untuk membersihkan, mengubah, dan memanipulasi data dengan berbagai operasi seperti pemilihan, penyaringan, pengelompokan, agregasi, dan penggabungan data. Ini memungkinkan analisis data untuk menjalankan tugas seperti eksplorasi data, pengolahan data, dan penyajian data dengan mudah.

Selain itu, Pandas menyediakan alat statistik dan visualisasi yang dapat membantu pengguna dalam melakukan analisis data lebih lanjut. Anda dapat dengan mudah menghitung statistik deskriptif, melakukan analisis regresi, dan membuat visualisasi data seperti grafik batang, diagram lingkaran, dan diagram sebar. Pandas juga bekerja dengan baik dengan alat visualisasi lainnya seperti Matplotlib dan Seaborn, sehingga memungkinkan pengguna untuk membuat visualisasi yang lebih menarik dan informatif.

Pandas telah menjadi alat yang sangat penting dalam komunitas ilmu data dan analisis data Python karena kemampuannya yang sangat fleksibel, dokumentasi yang kuat, dan komunitas pengguna yang besar. Framework ini memungkinkan pengguna untuk dengan mudah menjalankan berbagai tugas analisis data, dari pengolahan data dasar hingga analisis yang lebih kompleks, dan merupakan alat yang penting dalam ekosistem ilmu data Python.