

- **PCA**

Principal Component Analysis (PCA) adalah suatu metode statistik yang sangat berguna dalam analisis multivariat. Tujuannya adalah untuk mengatasi masalah dimensi tinggi dalam data dengan cara mengurangi dimensi data tersebut. PCA mencapai ini dengan mentransformasikan variabel-variabel asli ke dalam ruang dimensi yang lebih rendah, yang disebut sebagai komponen utama.

Langkah pertama dalam PCA adalah menghitung matriks kovariansi dari data asli. Matriks kovariansi ini mengukur hubungan antara variabel-variabel yang ada dalam data. Selanjutnya, 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 secara turun, kita dapat memilih sejumlah komponen utama tertentu yang akan digunakan untuk menggantikan data asli dalam dimensi yang lebih rendah. Hal ini membantu dalam mengidentifikasi pola-pola yang signifikan dalam data, serta mengurangi noise dan redundansi dalam data asli. Dengan demikian, PCA adalah alat yang kuat untuk analisis data yang kompleks.

- **LDA**

Linear Discriminant Analysis (LDA) adalah suatu teknik yang berfokus pada pemisahan kelas dalam analisis statistik dan pembelajaran mesin. Berbeda dengan PCA, yang hanya mempertimbangkan variasi total dalam data, LDA bertujuan untuk meningkatkan pemisahan kelas. Untuk mencapai tujuan ini, LDA bekerja dengan menghitung matriks scatter antar kelas dan matriks scatter dalam kelas dari data yang diberikan. Matriks scatter antar kelas mengukur sejauh mana data dari kelas yang berbeda tersebar, sementara matriks scatter dalam kelas mengukur variasi data dalam kelas yang sama.

Tujuan utama LDA adalah untuk mengoptimalkan pemisahan antar kelas sambil meminimalkan pemisahan dalam kelas. Hasil dari analisis LDA adalah vektor diskriminan atau komponen diskriminan. Komponen ini digunakan untuk mentransformasi data asli ke dalam dimensi yang lebih rendah, dengan tetap mempertahankan sebanyak mungkin informasi kelas. Dengan demikian, LDA adalah alat yang efektif untuk mengidentifikasi pola-pola kelas dalam data yang kompleks, dan ini sering digunakan dalam aplikasi klasifikasi, pengenalan pola, dan pemrosesan gambar.

- **SVD**

Singular Value Decomposition (SVD) adalah sebuah teknik dekomposisi matriks yang memiliki aplikasi luas di berbagai bidang, seperti matematika, statistik, dan ilmu data. SVD memecah matriks menjadi tiga komponen utama: matriks singular value, matriks left singular, dan matriks right singular. Dalam konteks analisis Principal Component Analysis (PCA), SVD digunakan untuk mendekomposisi matriks kovariansi dari data. Komponen utama dalam PCA dihasilkan dari matriks left singular. Ini memungkinkan kita untuk mengurutkan dan memilih komponen utama yang sesuai untuk reduksi dimensi data, membantu mengidentifikasi pola-pola penting dalam data asli.

Selain digunakan dalam PCA, SVD memiliki banyak aplikasi lainnya, termasuk reduksi dimensi, kompresi data, dan pemrosesan sinyal. Misalnya, SVD dapat digunakan untuk merekonstruksi data asli dari data yang telah direduksi dimensi. Hal ini memungkinkan kita untuk memahami data dengan lebih efisien dan mengurangi kehilangan informasi yang signifikan saat mengurangi dimensi. Oleh karena itu, SVD merupakan alat penting dalam analisis data dan pemrosesan informasi.

- **Pandas Framework**

Pandas merupakan sebuah framework open-source yang sangat populer dalam dunia analisis data dengan menggunakan bahasa pemrograman Python. Framework ini mempersembahkan alat yang sangat efektif dalam mengelola dan memanipulasi data tabular, mirip dengan format spreadsheet atau basis data SQL. Karena fleksibilitasnya, Pandas menjadi sangat berharga dalam berbagai aspek seperti analisis data, ilmu data, dan komputasi ilmiah secara umum. Pandas

menghadirkan dua struktur data utama yang menjadi landasan, yaitu Series dan DataFrame. Series adalah struktur data satu dimensi yang mirip dengan array atau list, sedangkan DataFrame adalah struktur data dua dimensi yang dapat diibaratkan sebagai tabel dengan baris dan kolom.

Pandas tidak hanya mempermudah pengguna dalam mengimpor data dari berbagai sumber, seperti file CSV, Excel, basis data SQL, dan bahkan web, tetapi juga menyediakan beragam operasi yang memungkinkan pengguna untuk membersihkan, mengubah, dan memanipulasi data dengan mudah. Dengan Pandas, pengguna dapat melakukan berbagai tugas analisis data seperti pemilihan, penyaringan, pengelompokan, agregasi, dan penggabungan data. Semua ini menjadikan Pandas sebagai alat yang kuat untuk menjalankan berbagai tugas dalam analisis data, mulai dari eksplorasi data hingga penyajian data dengan mudah dan efisien.