

Nama : Mauricio Bethoven Tigauw  
Nim : 1103204099  
Kelas : TK-44-G4

## PCA,LDA,SVD dan Pandas Framework

### Principal Component Analysis (PCA)

Principal Component Analysis (PCA) adalah teknik statistik yang digunakan untuk mengurangi dimensi data dengan tujuan untuk mengurangi kerumitan dan mempertahankan sebagian besar informasi yang ada dalam data. PCA mencari transformasi linier dari data asli ke ruang dimensi yang lebih rendah, di mana dimensi baru ini disusun dalam urutan menurun berdasarkan jumlah variasi dalam data. Dengan kata lain, PCA mengidentifikasi pola tersembunyi dalam data dan merepresentasikan data dalam bentuk dimensi yang lebih sedikit tanpa kehilangan banyak informasi.

Beberapa poin penting tentang PCA:

1. **Reduksi Dimensi:** PCA mengurangi dimensi dataset dengan menggantikan variabel-variabel awal dengan kombinasi linear baru dari variabel tersebut. Variabel baru ini disebut komponen utama atau principal components.
2. **Tujuan Utama:** Tujuan utama PCA adalah untuk mengurangi kerumitan dalam data sambil mempertahankan struktur dan informasi yang signifikan. Hal ini berguna terutama ketika data memiliki banyak variabel (dimensi tinggi) dan memungkinkan visualisasi data dalam bentuk yang lebih mudah dimengerti.
3. **Komponen Utama:** Komponen utama adalah vektor-vektor eigen dari matriks kovariansi dari data. Mereka menggambarkan arah di mana data bervariasi paling banyak. Komponen pertama adalah arah terpenting, diikuti oleh komponen kedua, dan seterusnya.
4. **Variansi:** PCA mengurutkan komponen utama berdasarkan jumlah varians yang dijelaskan oleh setiap komponen. Komponen utama pertama menjelaskan sebanyak mungkin varians dalam data, diikuti oleh komponen kedua yang menjelaskan sebanyak mungkin varians yang tersisa, dan seterusnya.
5. **Keuntungan:** Selain pengurangan dimensi, PCA juga berguna untuk mengatasi masalah multikolinearitas (ketika variabel-variabel dalam data berkorelasi tinggi satu sama lain) dan dapat meningkatkan kinerja model-machine learning dengan mengurangi kelebihan variabilitas dalam data.
6. **Aplikasi:** PCA banyak digunakan dalam bidang seperti pengolahan citra, pengenalan pola, analisis biomedis, analisis spektral, dan berbagai bidang ilmu lainnya di mana data memiliki dimensi yang tinggi.

Dalam praktiknya, PCA melibatkan perhitungan nilai eigen dari matriks kovariansi dari data atau matriks korelasi (tergantung pada konteks), diikuti oleh proyeksi data ke ruang yang lebih rendah

menggunakan komponen utama yang dihasilkan. PCA adalah alat yang sangat berguna untuk memahami struktur dalam data dan mengatasi masalah dengan dimensi tinggi.

### **Linear Discriminant Analysis (LDA)**

Linear Discriminant Analysis (LDA) adalah teknik statistik yang digunakan untuk pemisahan linier dari dua atau lebih kelas atau kelompok dalam sebuah dataset. Tujuan utama LDA adalah mengidentifikasi kombinasi linear dari variabel-variabel input yang paling baik membedakan antara dua atau lebih kelas dalam dataset.

Berbeda dengan Principal Component Analysis (PCA), yang mengurangi dimensi data tanpa mempertimbangkan label kelas, LDA melakukan reduksi dimensi dengan mempertimbangkan label kelas variabel input. Oleh karena itu, LDA adalah metode pengawasan (supervised learning).

Beberapa poin penting tentang LDA:

1. **Maksimalkan Jarak Antar Kelas:** LDA bertujuan untuk membangun proyeksi baru (komponen atau linear discriminants) sedemikian rupa sehingga jarak antara pusat-pusat kelas (rata-rata kelas) maksimal dan variabilitas dalam setiap kelas (dispersi dalam kelas) minimal.
2. **Transformasi Linear:** LDA melibatkan transformasi linear dari variabel input ke dalam bentuk kombinasi linear baru yang memaksimalkan pemisahan antara kelas. Transformasi ini didasarkan pada matriks kovariansi dan vektor rata-rata dari setiap kelas.
3. **Proyeksi Data:** Setelah transformasi dilakukan, data dapat diproyeksikan ke ruang dimensi yang lebih rendah (biasanya satu dimensi jika hanya ada dua kelas) untuk klasifikasi atau analisis lebih lanjut.
4. **Keuntungan LDA:** LDA membantu mempertahankan sebanyak mungkin informasi kelas dalam proyeksi linier, yang membuatnya bermanfaat dalam pengklasifikasian dan memahami perbedaan antara kelas-kelas dalam data.
5. **Aplikasi:** LDA sering digunakan dalam pengenalan pola, pengolahan citra, dan bidang-bidang di mana klasifikasi berbasis pada kombinasi linear dari fitur-fitur adalah relevan.

Pada dasarnya, LDA membantu mengatasi masalah klasifikasi ketika ada dua atau lebih kelas dan variabel-variabel input. Dengan menemukan kombinasi linear dari variabel-variabel tersebut, LDA memungkinkan pemisahan yang lebih baik antara kelas-kelas ini, membuatnya menjadi alat yang berguna dalam analisis data dan pengambilan keputusan.

### **Singular Value Decomposition (SVD)**

Singular Value Decomposition (SVD) adalah teknik matriks faktorisasi yang digunakan dalam aljabar linear dan analisis numerik. SVD menguraikan suatu matriks menjadi tiga matriks lebih sederhana, memungkinkan pemahaman yang lebih baik tentang struktur matriks dan memfasilitasi berbagai operasi

pada data. SVD berguna dalam berbagai aplikasi, termasuk reduksi dimensi, analisis faktor, kompresi data, dan pemulihan informasi dari data yang terkontaminasi atau tidak lengkap.

Untuk suatu matriks  $A$  dengan  $m$  baris dan  $n$  kolom, SVD menguraikan  $A$  sebagai hasil perkalian tiga matriks:

$$A = U\Sigma V^T$$

di mana:

- $U$  adalah matriks ortogonal  $m \times m$  (kolom-kolomnya adalah vektor-vektor eigen dari  $AA^T$ ).
- $\Sigma$  adalah matriks diagonal  $m \times n$  yang memiliki nilai singular values  $\sigma_1, \sigma_2, \dots, \sigma_r$  (diurutkan dalam urutan menurun) di diagonal utama, diikuti oleh nol.
- $V^T$  adalah transpose dari matriks ortogonal  $n \times n$  (kolom-kolomnya adalah vektor-vektor eigen dari  $AA^T$ ).

Penting untuk dicatat bahwa SVD selalu mungkin dilakukan untuk setiap matriks  $A$ , meskipun  $A$  bukanlah matriks persegi.

Aplikasi SVD meliputi:

1. **Reduksi Dimensi:** Dalam analisis data, SVD digunakan untuk mengurangi dimensi dataset dengan menghapus singular values yang kecil, memungkinkan representasi yang lebih efisien dari data.
2. **Kompresi Data:** Dengan mempertahankan hanya beberapa singular values terbesar, data dapat dikompresi tanpa kehilangan banyak informasi.
3. **Pemulihan Informasi:** SVD dapat digunakan untuk memperkirakan nilai-nilai yang hilang dalam data terkontaminasi atau data yang tidak lengkap.
4. **Analisis Faktor:** Dalam statistika, SVD dapat digunakan untuk analisis faktor, yang melibatkan ekstraksi pola dalam data.
5. **Pemrosesan Gambar dan Suara:** SVD digunakan dalam berbagai aplikasi pemrosesan gambar dan suara, seperti kompresi gambar (seperti dalam format JPEG) dan pengolahan sinyal suara.

SVD memiliki banyak aplikasi dan penting dalam berbagai bidang ilmu pengetahuan dan teknologi karena kemampuannya untuk mengungkapkan struktur matriks yang kompleks dan menyederhanakannya menjadi komponen-komponen yang lebih sederhana.

Pandas adalah salah satu perpustakaan (library) Python yang populer untuk analisis data. Pandas menyediakan struktur data dan fungsi manipulasi data yang mudah digunakan, memungkinkan pengguna untuk melakukan tugas-tugas seperti pembacaan dan penulisan data dalam berbagai format, pengindeksan, pemfilteran, penggabungan, dan transformasi data dengan mudah.

Beberapa komponen kunci dari Pandas adalah:

1. **DataFrame:** DataFrame adalah struktur data dua dimensi mirip tabel yang terdiri dari kolom-kolom dengan label dan baris-baris dengan label. DataFrame memungkinkan pengguna untuk

menyimpan dan memanipulasi data dengan cara yang mirip dengan spreadsheet atau tabel database SQL.

2. **Series:** Series adalah struktur data satu dimensi yang dapat menyimpan data dari berbagai jenis, termasuk integer, float, dan string. Series memiliki indeks yang dapat diatur oleh pengguna.
3. **Pembacaan dan Penulisan Data:** Pandas mendukung berbagai format file, termasuk CSV, Excel, SQL, JSON, dan banyak lagi. Pengguna dapat membaca data dari file dan menyimpan hasil analisis kembali ke berbagai format tersebut.
4. **Indeks dan Seleksi Data:** Pandas memungkinkan pengguna untuk mengindeks data berdasarkan nama kolom dan/atau indeks baris. Ini memungkinkan akses dan manipulasi data dengan cara yang sangat fleksibel.
5. **Manipulasi Data:** Pandas menyediakan berbagai fungsi dan metode untuk membersihkan, memfilter, menggabungkan, menghitung statistik, dan melakukan operasi lainnya pada data.
6. **Manajemen Missing Data:** Pandas menyediakan alat untuk mengatasi data yang hilang (missing data) dengan menghapusnya atau menggantinya dengan nilai tertentu.
7. **Operasi Grouping:** Pandas mendukung operasi grouping data yang mirip dengan SQL, memungkinkan pengguna untuk menerapkan fungsi agregasi pada grup data berdasarkan kriteria tertentu.

Pandas sangat penting dalam analisis data dan ilmu data karena kemudahan penggunaannya dan fungsionalitasnya yang kuat. Dengan menggunakan Pandas, pengguna dapat dengan mudah mengimpor, memanipulasi, dan menganalisis data dalam bahasa Python, menjadikannya salah satu alat pilihan untuk ilmuwan data, analis data, dan pengembang Python dalam bekerja dengan data.