

Nama : Al Ghifary Akmal Nasheeri

NIM : 1103201242

Kelas : TK-44-06

PCA (Principal Component Analysis)

PCA adalah metode dimensionality reduction yang digunakan untuk mengurangi dimensi data tanpa kehilangan terlalu banyak informasi. PCA bekerja dengan menemukan komponen utama dari data, yaitu arah-arah varians tertinggi. Komponen-komponen utama ini kemudian digunakan untuk merepresentasikan data dalam dimensi yang lebih rendah.

Penjelasan lebih detail mengenai PCA adalah sebagai berikut:

- **Data dimensional tinggi**

Data dimensional tinggi adalah data yang memiliki banyak fitur. Misalnya, data gambar memiliki tiga fitur (merah, hijau, dan biru) untuk setiap piksel, sehingga datanya memiliki dimensi $3 * 1000 * 1000 = 3,000,000$. Data dimensional tinggi dapat sulit untuk dianalisis dan diinterpretasi.

- **Komponen utama**

Komponen utama adalah arah-arah varians tertinggi dalam data. Varians adalah ukuran seberapa tersebar data. Komponen utama adalah arah-arah yang dapat menjelaskan sebanyak mungkin varians dalam data dengan jumlah dimensi yang paling sedikit.

- **Proyeksi data**

Data dapat diproyeksikan ke ruang dimensi yang lebih rendah dengan menggunakan komponen utama. Proyeksi ini dilakukan dengan mengalikan data dengan matriks transformasi yang dihitung dari komponen utama.

Contoh:

Misalkan kita memiliki data gambar dengan tiga fitur (merah, hijau, dan biru) untuk setiap piksel. Kita dapat menggunakan PCA untuk mengurangi dimensi data dari 3 menjadi 2. Komponen utama pertama akan menjelaskan sebanyak mungkin varians dalam data, dan komponen utama kedua akan menjelaskan sisa varians.

Setelah data diproyeksikan ke ruang dimensi yang lebih rendah, kita dapat menggunakan visualisasi untuk menganalisis data. Misalnya, kita dapat menggunakan scatter plot untuk melihat bagaimana data dikelompokkan dalam dimensi yang lebih rendah.

LDA (Linear Discriminant Analysis)

LDA adalah metode dimensionality reduction yang digunakan untuk mengurangi dimensi data sambil mempertahankan informasi yang relevan untuk klasifikasi. LDA bekerja dengan menemukan proyeksi data ke ruang dimensi yang lebih rendah di mana kelas-kelas berbeda dipisahkan secara maksimal.

Penjelasan lebih detail mengenai LDA adalah sebagai berikut:

- **Data berlabel**

LDA dirancang untuk digunakan dengan data berlabel, yaitu data yang memiliki kelas yang diketahui untuk setiap pengamatan.

- **Proyeksi data**

Data diproyeksikan ke ruang dimensi yang lebih rendah dengan menggunakan matriks transformasi yang dihitung dari data dan kelas. Matriks transformasi ini dirancang untuk memisahkan kelas-kelas secara maksimal dalam dimensi yang lebih rendah.

Contoh:

Misalkan kita memiliki data gambar bunga dengan dua kelas: mawar dan tulip. Kita dapat menggunakan LDA untuk mengurangi dimensi data dari 3 menjadi 2. Matriks transformasi akan dihitung dari data dan kelas.

Setelah data diproyeksikan ke ruang dimensi yang lebih rendah, kita dapat menggunakan visualisasi untuk melihat bagaimana kelas-kelas dipisahkan. Misalnya, kita dapat menggunakan scatter plot untuk melihat bagaimana data dikelompokkan dalam dimensi yang lebih rendah.

SVD (Singular Value Decomposition)

SVD adalah metode untuk mendekomposisi matriks menjadi produk dari tiga matriks: matriks U, matriks S, dan matriks V. SVD memiliki banyak aplikasi, termasuk dimensionality reduction, filtering, dan rekomendasi sistem.

Penjelasan lebih detail mengenai SVD adalah sebagai berikut:

- **Dekomposisi matriks**

SVD mendekomposisi matriks menjadi produk dari tiga matriks:

$$A = U * S * V^T$$

- **Matriks U**

Matriks U adalah matriks ortogonal yang merepresentasikan vektor-vektor dasar dari ruang eigen dari matriks A.

- **Matriks S**

Matriks S adalah matriks diagonal yang berisi nilai-nilai eigen dari matriks A .

- **Matriks V^T**

Matriks V^T adalah transpose dari matriks V .

SVD memiliki banyak aplikasi, termasuk:

- **Dimensionality reduction**

SVD dapat digunakan untuk mengurangi dimensi data dengan menggunakan komponen utama dari matriks S .

- **Filtering**

SVD dapat digunakan untuk memfilter noise dari data dengan menggunakan nilai-nilai eigen dari matriks S .

- **Rekomendasi sistem**

SVD dapat digunakan untuk merekomendasikan produk kepada pengguna berdasarkan riwayat pembelian mereka.

Pandas framework

Pandas adalah Python library untuk analisis dan manipulasi data. Pandas menyediakan struktur data yang fleksibel dan efisien untuk menyimpan dan mengelola data, serta berbagai fungsi untuk melakukan operasi pada data.

Penjelasan lebih detail:

- **Struktur data**

Pandas menyediakan dua struktur data utama:

- **DataFrame**

DataFrame adalah struktur data yang mirip dengan spreadsheet. DataFrame terdiri dari baris dan kolom, dan setiap sel berisi data.

- **Series**

Series adalah struktur data satu dimensi yang mirip dengan array. Series terdiri dari urutan data, dan setiap data memiliki label.

- **Operasi data**

Pandas menyediakan berbagai fungsi untuk melakukan operasi data, seperti:

- **Membaca dan menulis data**

Pandas dapat digunakan untuk membaca dan menulis data dari berbagai sumber, seperti file CSV, Excel, dan database.

- **Membersihkan dan preprocessing data**

Pandas dapat digunakan untuk membersihkan dan preprocessing data, seperti menghapus data yang hilang, menangani data yang tidak valid, dan mengubah format data.

- **Analisis data**

Pandas dapat digunakan untuk melakukan analisis data, seperti visualisasi data, statistik deskriptif, dan analisis regresi.