

Nama : Muhammad Daffa

Kelas : TK-44-G4

NIM : 1103201258

1. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) adalah salah satu metode reduksi dimensi pada machine learning. PCA akan memilih variabel-variabel yang mampu menjelaskan sebagian besar variabilitas data. PCA mengurangi dimensi dengan membentuk variabel-variabel baru yang disebut Principal Components. Principal Components yang merupakan kombinasi linier dari variabel-variabel lama. Penghitungan Varians dan Principal Component ini dapat dilakukan dengan menggunakan konsep nilai eigen (eigenvalue) dan vektor eigen (eigenvector) dari ilmu Aljabar Linier.

Principal Component Analysis (PCA) tentunya punya banyak manfaat dalam proses analisis data, misal:

- Mengatasi multikolinieritas yang pada metode parametrik tertentu merupakan asumsi yang harus dipenuhi;
- Mereduksi jumlah variabel yang akan dimasukkan ke model;
- Jumlah variabel yang lebih sedikit tentu akan menyederhanakan model;
- Juga mempercepat proses komputasi.

Principal Component Analysis (PCA) adalah metode reduksi yang populer. Sehingga telah banyak diimplementasikan ke bahasa pemrograman maupun perangkat lunak pengolah data. Salah satunya adalah bahasa R. Pada artikel ini, kita akan mempraktekkan materi pada modul Data Science in Finance: Dimension Reduction.

2. Linear Discriminant Analysis

Analisis Diskriminan Linear (Linear Discriminant Analysis atau LDA) adalah sebuah metode statistik yang digunakan dalam analisis multivariat dan pembelajaran mesin. Tujuan utama dari LDA adalah untuk memproyeksikan data dari beberapa kelas (kelompok) ke sebuah ruang dimensi yang lebih rendah sehingga perbedaan antara kelas-kelas tersebut menjadi lebih terlihat.

Berikut adalah ringkasan tentang Linear Discriminant Analysis:

Tujuan Utama:

LDA digunakan untuk pemisahan dan klasifikasi data ke dalam beberapa kelas atau kelompok berdasarkan informasi dalam data tersebut. Tujuannya adalah untuk menciptakan proyeksi atau transformasi data ke dalam dimensi yang lebih rendah, sehingga perbedaan antara kelas-kelas menjadi lebih jelas.

Proses Utama:

LDA memulai dengan menghitung matriks dispersi dalam kelas (within-class scatter matrix) dan matriks dispersi antar kelas (between-class scatter matrix).

Selanjutnya, LDA mencari vektor proyeksi yang disebut vektor diskriminan linear, yang dapat memaksimalkan rasio antara matriks dispersi antar kelas dan matriks dispersi dalam kelas.

Data diproyeksikan ke ruang yang lebih rendah dengan menggunakan vektor diskriminan linear ini.

Keuntungan:

LDA membantu dalam memahami perbedaan antara kelas atau kelompok data, sehingga memudahkan tugas klasifikasi.

LDA mengurangi dimensi data, yang dapat mengurangi masalah kutukan dimensi (curse of dimensionality) dan memperbaiki kinerja model.

Kelemahan:

LDA bersifat supervisi, yang berarti kita memerlukan label kelas pada data pelatihan untuk melaksanakan analisis ini.

LDA tidak efektif jika asumsi dasar distribusi data tidak terpenuhi, seperti asumsi normalitas.

Aplikasi:

LDA digunakan dalam berbagai bidang, termasuk pengenalan pola, pengolahan citra, biometrik, analisis ekonomi, dan pemrosesan bahasa alami.

Ini juga digunakan dalam pengenalan wajah, pengenalan tulisan tangan, dan pemrosesan teks.

LDA adalah alat yang kuat dalam pemahaman dan klasifikasi data dalam beberapa kelas berbeda. Dengan cara yang serupa, ada juga analisis diskriminan kuadrat (Quadratic Discriminant Analysis atau QDA), yang memiliki asumsi distribusi data yang berbeda dan dapat digunakan dalam situasi yang berbeda.

3. Singular Value Decomposition

Singular Value Decomposition (SVD) adalah teknik dekomposisi matriks yang sangat berguna dalam berbagai aplikasi, terutama dalam analisis data, pemrosesan sinyal, dan pembelajaran mesin. SVD menguraikan matriks menjadi tiga matriks yang lebih sederhana, yaitu U , Σ , dan V :

- Matriks U adalah matriks orthogonal ($U^T * U = I$) yang berisi vektor-vektor eigen untuk matriks asli AA^T (matriks kovariansi).
- Matriks Σ adalah matriks diagonal yang berisi singular values dalam urutan menurun. Singular values adalah akar positif dari eigenvalues matriks kovariansi. Matriks Σ mengukur pentingnya setiap mode (komponen) dalam data.
- Matriks V adalah matriks orthogonal yang berisi vektor-vektor eigen untuk matriks asli A^TA (matriks kovariansi alternatif).

Dengan SVD, kita dapat melakukan reduksi dimensi, kompresi data, dan pemrosesan matriks yang lebih efisien. Beberapa aplikasi SVD meliputi:

- Principal Component Analysis (PCA): PCA menggunakan SVD untuk mengekstrak komponen utama dalam data dan mengurangi dimensi data.
- Rekonstruksi Matriks: SVD dapat digunakan untuk mengestimasi matriks asli dari dekomposisi, yang berguna dalam berbagai aplikasi termasuk kompresi gambar dan pemrosesan sinyal.
- Pemutusan Matriks: SVD digunakan dalam faktorisasi matriks untuk merekonstruksi data yang hilang atau dalam collaborative filtering dalam sistem rekomendasi.
- Pemrosesan Sinyal: SVD digunakan untuk analisis sinyal audio dan gambar, seperti pengurangan noise dan kompresi.

SVD adalah alat matematika yang kuat dan serbaguna dalam analisis data dan pengolahan informasi.