

Understanding PCA, LDA, SVD on Machine learning

Fahmi Nanda Saputro_1103200203

1. PCA

Principal Component Analysis

■ PCA is “an orthogonal linear transformation that transfers the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (first principal component), the second greatest variance lies on the second coordinate (second principal component), and so on.”

- An exploratory technique to reduce the dimensionality of the data set, used to:
 1. Reduce number of dimensions in data
 2. Find patterns in high-dimensional data
 3. Visualize data of high dimensionality

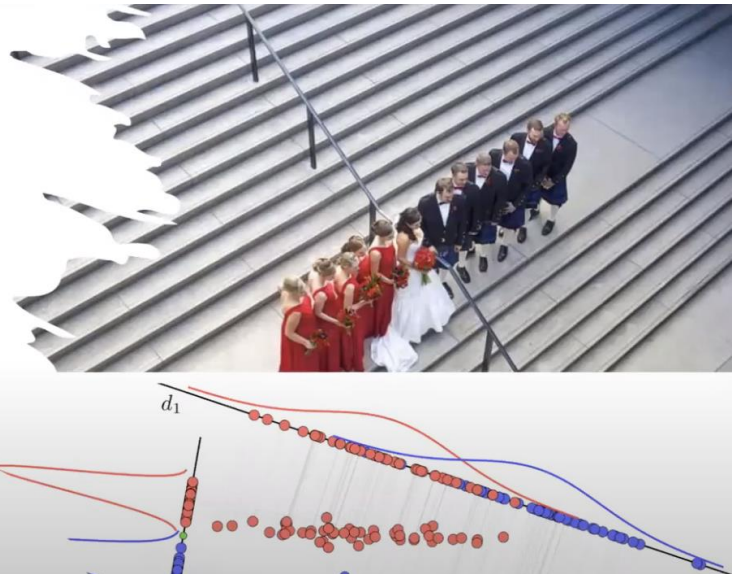
Merupakan metode statistik dan Teknik reduksi dimensi yang biasa digunakan dalam machine learning, dengan memproyeksikan data dari ruang dimensi tinggi ke rendah dengan mempertahankan sebanyak mungkin informasi yang relevan.

Terdapat, reduksi dimensi untuk mengurangi jumlah fitur atau atribut, komponen utama PCA adalah mengidentifikasi komponen utama atau vector yang menggambarkan variasi paling signifikan dalam data. Ortogonal merupakan sifat komponen utama dalam PCA untuk memisahkan variasi data dalam reduksi dimensi.

2. LDA

Linear Discriminant Analysis

- **LDA:** Pick a new dimension that gives:
 - Maximum separation of means of projected classes
 - Minimum variance between each projected class
- **Solution:** Eigenvectors based on between-class and within-class covariance matrices.



Merupakan Teknik statistik yang digunakan untuk mengurangi dimensi data sambil mempertahankan informasi paling relevan untuk pemisahan kelas atau kategori data. Pemisahan kelas di LDA digunakan untuk meningkatkan pemisahan antara berbagai

kelas. Reduksi Dimensi digunakan di LDA untuk mengurangi dimensi data, tetapi dengan penekanan pada pemisahan data. Proses perhitungan matrix dalam LDA melibatkan Matriks kovarian dalam antar kelas dan menghitung vector eigen.

3.

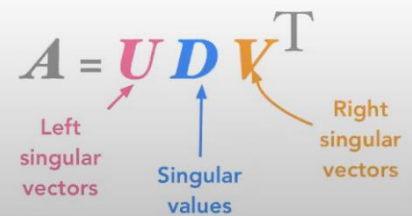
Singular Value Decomposition

- Let A be an $m \times n$ matrix. The Singular Value Decomposition(SVD) of A :

$$A = U D V^T$$

- U is orthogonal $m \times r$, V is $n \times r$ orthogonal and D is $r \times r$ diagonal matrices.

- A : Input data matrix (eg. m documents, n terms)
- U : Left singular matrix (m documents, r **concepts**)
- D : Singular value (strength of each **concept**)
- V : Right singular matrix (n terms, r **concepts**)



Merupakan Teknik matematis yang digunakan untuk menguraikan matriks menjadi tiga matriks yang lebih sederhana, terdapat matriks input yang mungkin berisi nilai – nilai dari dataset. Decompsition berisi 3 matriks yang terdiri dari Matriks U, Sigma, dan V. SVD adalah alat yang kuat dalam analisis data dan machine learning karena dapat membantu mengidentifikasi struktur dalam data dan mengurangi dimensi data dengan mempertahankan informasi yang paling penting.