

## **StatQuest – Understanding PCA, LDA, SVD**

### **1. PCA**

Principal Component Analysis (PCA) adalah metode statistik yang membantu mengurangi kompleksitas data dengan mengubahnya menjadi bentuk yang lebih sederhana. Teknik ini memungkinkan identifikasi pola dan hubungan antar variabel. Tujuan dari PCA adalah untuk menemukan kombinasi linear dari variabel yang menangkap sebagian besar variasi data, sehingga menghasilkan kumpulan data yang lebih ringkas namun tetap mempertahankan informasi yang berharga. Proses PCA melibatkan transformasi data ke dalam sistem koordinat baru yang dikenal sebagai "komponen utama".

Komponen-komponen ini mewakili arah di mana variasi data paling besar. PC1 adalah komponen utama, yang menjelaskan variasi paling banyak, diikuti oleh PC2, PC3, dan seterusnya. PCA memastikan bahwa PC2 tegak lurus dengan PC1, PC3 tegak lurus dengan PC1 dan PC2, dan seterusnya. Setiap komponen utama memiliki "eigenvalue" yang mengukur kemampuannya untuk menjelaskan variasi data. Selain itu, "eigenvector" menunjukkan arah dari setiap komponen utama, sedangkan "nilai muatan" menunjukkan kontribusi setiap variabel terhadap komponen utama.

### **2. LDA**

Linear Discriminant Analysis (LDA) adalah metode statistik dan teknik machine learning yang digunakan untuk analisis data dan pengenalan pola. Tujuan utamanya adalah untuk mengidentifikasi perbedaan antara kelompok atau kelas dalam data dengan menemukan kombinasi linear dari variabel yang dapat memisahkan kelompok secara efektif. LDA sangat berguna dalam masalah klasifikasi yang melibatkan banyak kelas. Proses LDA melibatkan transformasi data ke dalam ruang berdimensi lebih rendah yang disebut "ruang diskriminan".

Transformasi ini membantu membuat perbedaan antara kelompok-kelompok menjadi lebih jelas dengan mengelompokkan data di dalam setiap kelompok dan memisahkan data di antara kelompok-kelompok tersebut. Dalam ruang diskriminan, LDA mencari "garis diskriminan" atau "hyperplane" yang optimal untuk mencapai pemisahan ini. Selain itu, LDA menghitung nilai eigen dan vektor eigen untuk mengukur sejauh mana perbedaan antar kelompok dijelaskan oleh garis diskriminan. Nilai eigen mengukur jumlah varians yang dijelaskan oleh setiap komponen linier, sedangkan vektor eigen menggambarkan arah garis diskriminan.

### **3. SVD**

SVD, atau Dekomposisi Nilai Singular, adalah teknik matematika yang digunakan dalam analisis matriks dan komputasi numerik. SVD sering digunakan dalam berbagai aplikasi, termasuk analisis data, pemrosesan gambar, kompresi data, dan banyak lagi. SVD memiliki banyak aplikasi yang berguna, terutama dalam analisis data dan pengenalan pola.

Proses SVD melibatkan penguraian atau pemisahan matriks asal menjadi tiga matriks terpisah:  $U$ ,  $\Sigma$ , dan  $V$ . Berikut ini adalah penjelasan lebih rinci tentang ketiga matriks tersebut:

1. Matriks  $U$  adalah matriks "vektor singular" kiri, yang berisi vektor-vektor yang mendefinisikan transformasi linear antara ruang asal dan ruang target.
2. Matriks  $\Sigma$  adalah matriks diagonal yang berisi "nilai singular". Nilai-nilai singular ini mengukur tingkat variasi atau pentingnya setiap komponen dalam data. Nilai-nilai tersebut diurutkan dalam urutan menurun, sehingga nilai singular pertama memiliki kontribusi paling besar dalam menggambarkan data.
4. Matriks  $V$  adalah matriks "vektor singular" sebelah kanan, yang berisi vektor-vektor yang menggambarkan transformasi linear antara ruang target dan ruang asal.

SVD sangat berguna dalam berbagai konteks, termasuk pengurangan dimensi data, kompresi data, pemrosesan gambar, dan pemrosesan teks. Salah satu aplikasi SVD yang paling umum adalah dalam PCA (Analisis Komponen Utama), di mana SVD digunakan untuk menghitung komponen utama data.