

Homework

- StatQuests ([PCA](#), [LDA](#), [SVD](#))
- Learn [PCA](#), [LDA](#) and [SVD](#) in ScikitLearn
- Learn Pandas Framework

Analisis Data Eksploratif (EDA) adalah proses yang digunakan dalam ilmu data untuk mengeksplorasi dan menganalisis data sebelum melakukan analisis statistik lebih lanjut atau membangun model prediksi. Tujuan utama dari EDA adalah untuk memahami sifat-sifat dasar dari dataset, mengidentifikasi pola, anomali, dan tren, serta membuat hipotesis awal tentang data sebelum memasuki tahap analisis yang lebih mendalam.

Exploratory Data Analysis

John Tukey (1961) EDA Focus on:

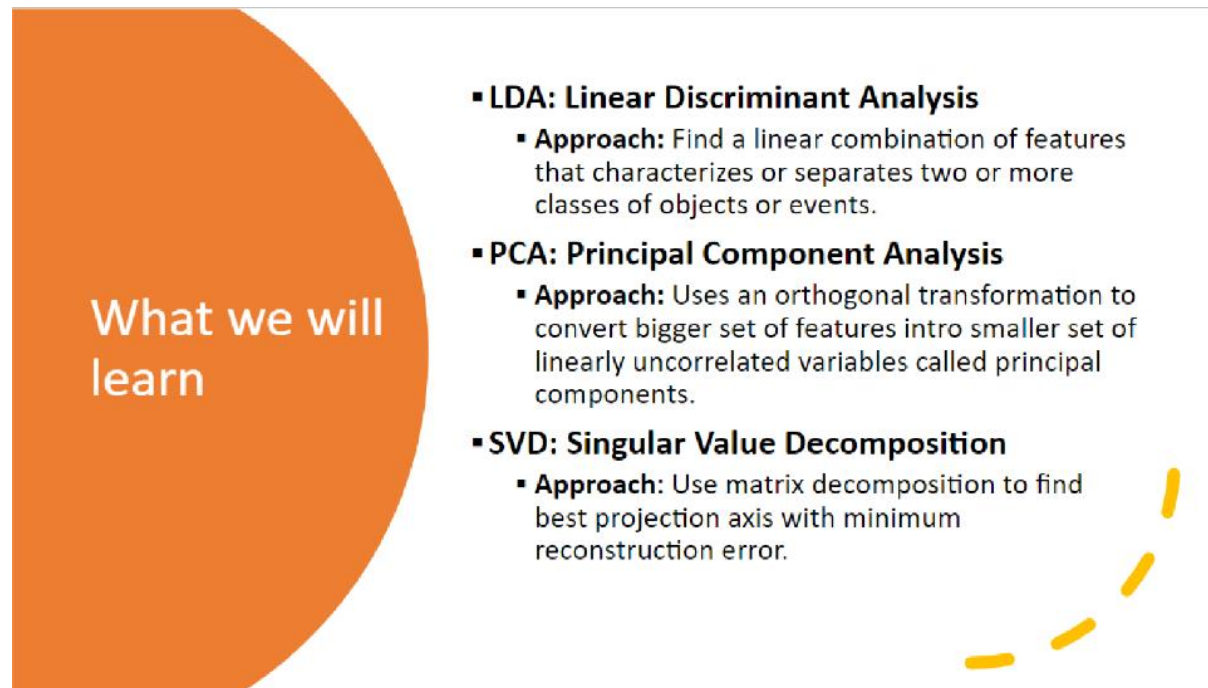
- Understand the data's **underlying structure**
- Develop **intuition** about the data set
- How the data were collected (to aid in **cleaning**)
- How to further **investigate** with statistical methods

EDA is Any Initial Investigations of Data:

1. **Basic:** Data Visualization and Numerical Summary
2. **Intermediate:** Statistical Hypothesis Tests with Confirmatory Analysis
3. **Advanced:** Unsupervised learning, PCA, SVD and Clustering



John Tukey, seorang ilmuwan statistik terkemuka, adalah salah satu tokoh utama dalam pengembangan Analisis Data Eksploratif (*Exploratory Data Analysis* - EDA). EDA adalah pendekatan dalam statistik yang diperkenalkan oleh Tukey pada tahun 1970-an yang berfokus pada pemahaman dan eksplorasi data sebelum melakukan analisis statistik formal. Tukey berpendapat bahwa EDA adalah langkah penting dalam pemahaman data sebelum melangkah ke analisis yang lebih mendalam.



What we will learn

- **LDA: Linear Discriminant Analysis**
 - **Approach:** Find a linear combination of features that characterizes or separates two or more classes of objects or events.
- **PCA: Principal Component Analysis**
 - **Approach:** Uses an orthogonal transformation to convert bigger set of features into smaller set of linearly uncorrelated variables called principal components.
- **SVD: Singular Value Decomposition**
 - **Approach:** Use matrix decomposition to find best projection axis with minimum reconstruction error.

LDA (*Linear Discriminant Analysis*) merupakan metode statistik dalam analisis multivariat. LDA digunakan untuk pemilihan fitur dan pemisahan kelas dalam data. Ini adalah teknik yang digunakan dalam pembelajaran mesin untuk pengurangan dimensi dan analisis statistik.

PCA adalah teknik pengurangan dimensi yang digunakan untuk menganalisis dan mengurangi kompleksitas dalam data. Teknik ini sering digunakan dalam analisis data multivariat dan dalam pemrosesan citra. Tujuan utama dari PCA adalah mengidentifikasi komponen-komponen utama (*principal components*) dalam data yang menjelaskan sebagian besar variasi dalam data, sehingga dimungkinkan untuk merepresentasikan data dalam dimensi yang lebih rendah daripada dimensi aslinya, tetapi tetap mempertahankan informasi yang signifikan.

SVD adalah singkatan dari "*Singular Value Decomposition*" (Pemisahan Nilai Tunggal) dalam matematika dan analisis data. SVD adalah teknik dekomposisi matriks yang digunakan untuk berbagai tujuan dalam analisis data, pemrosesan sinyal, pengenalan pola, dan banyak aplikasi lainnya. SVD adalah alat yang kuat dalam analisis data dan pemrosesan sinyal, dan sering digunakan dalam berbagai aplikasi ilmu data dan teknik lainnya.