Introduction to Principal Component Analysis

What have we covered till now?

- Feature Selection/Reduction
  - Missing value ratio
  - Low variance
  - High correlation
  - Backward feature elimination
  - Forward feature selection

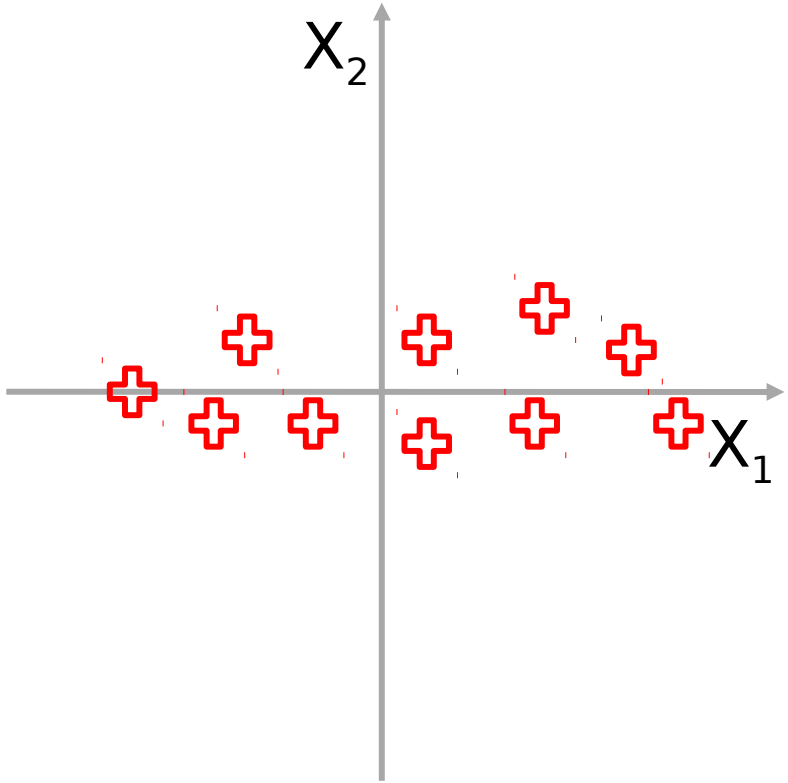
What have we covered till now?

- Feature Selection/Reduction
  - Missing value ratio
  - Low variance
  - High correlation
  - Backward feature elimination
  - Forward feature selection
- Advanced Feature Reduction Techniques
  - Principal Component Analysis (PCA)
  - Factor Analysis

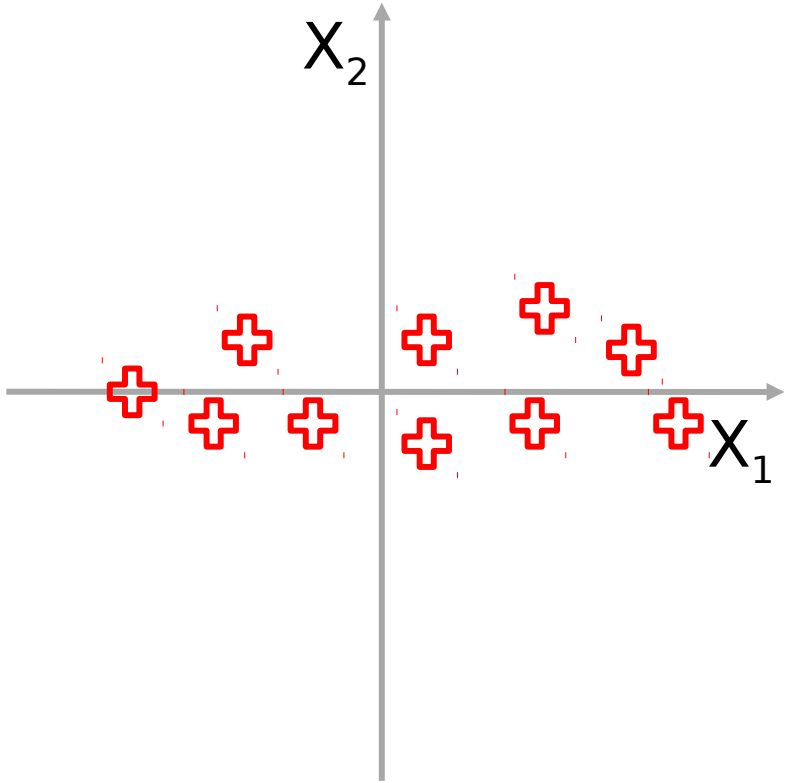
What are Principal Components?

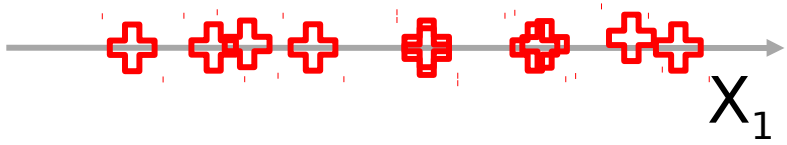
The 2-Dimensional Case

The 2-Dimensional Case



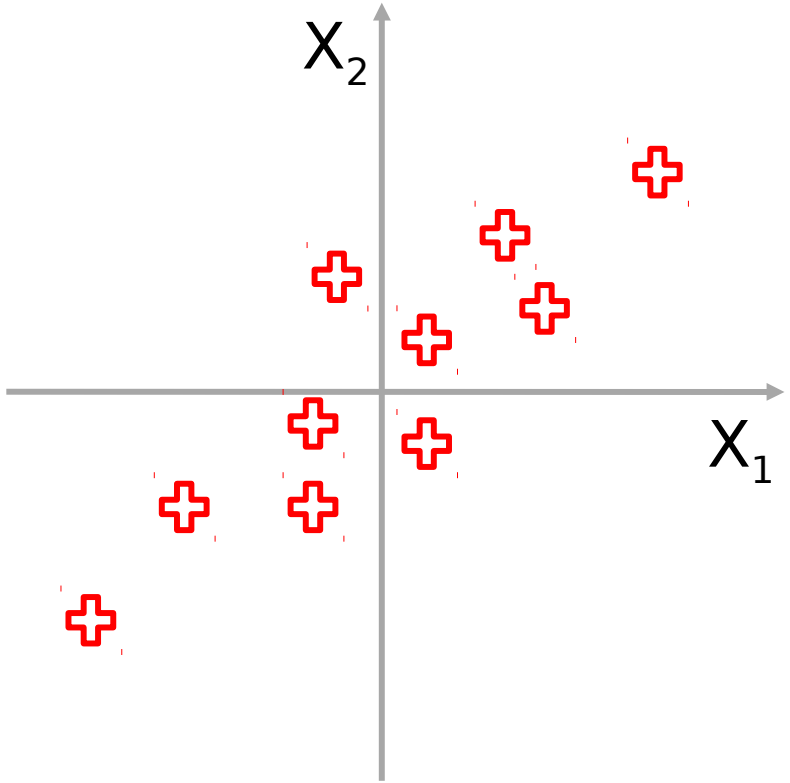
The 2-Dimensional Case





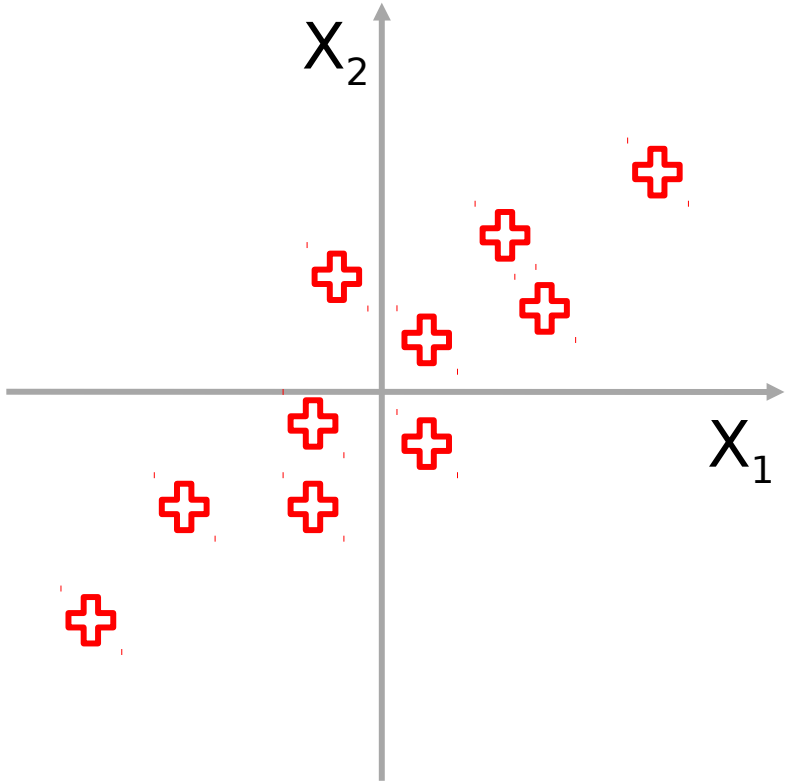
Analytics Vidhya
Learn everything about analytics

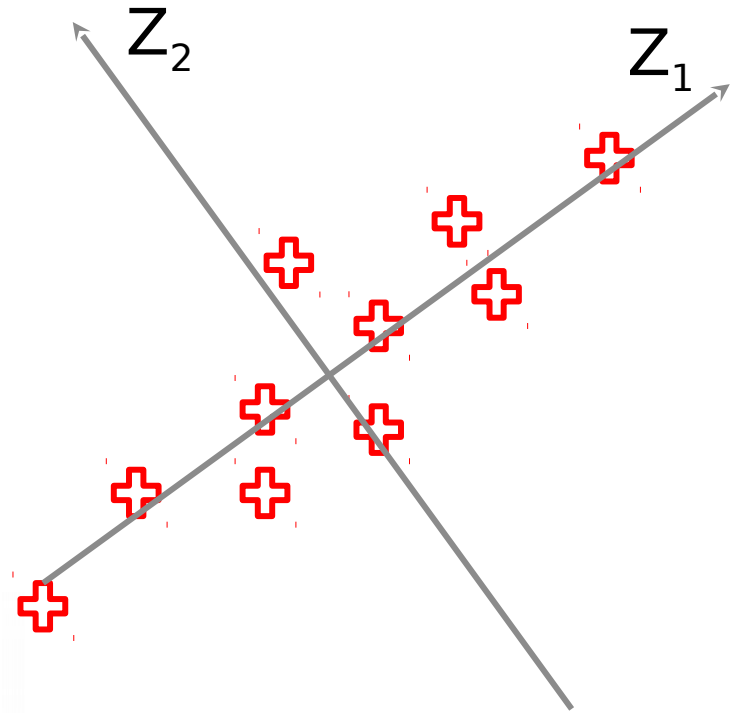
The 2-Dimensional Case



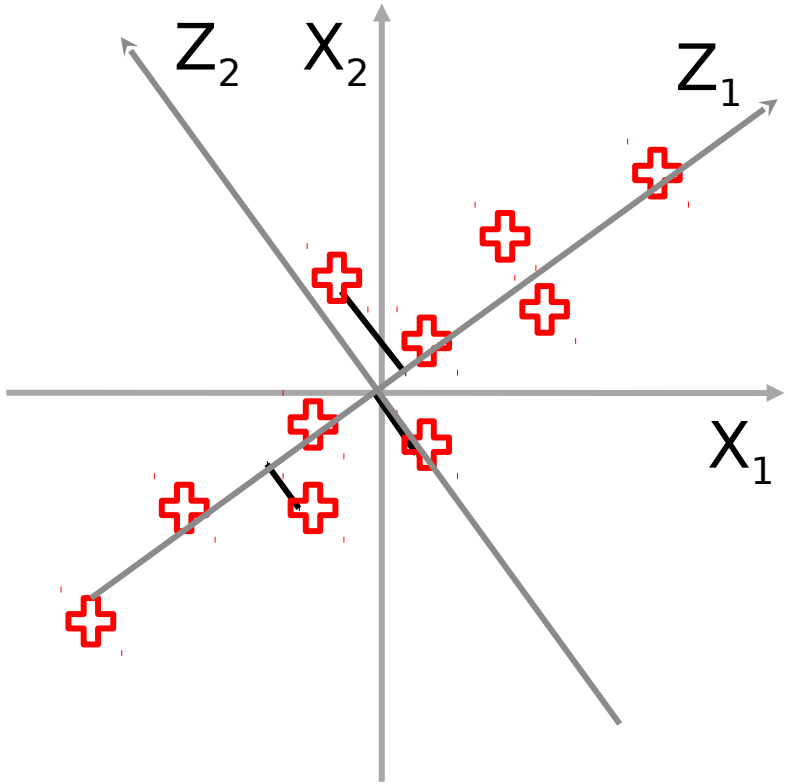
None of the features has low variance

The 2-Dimensional Case





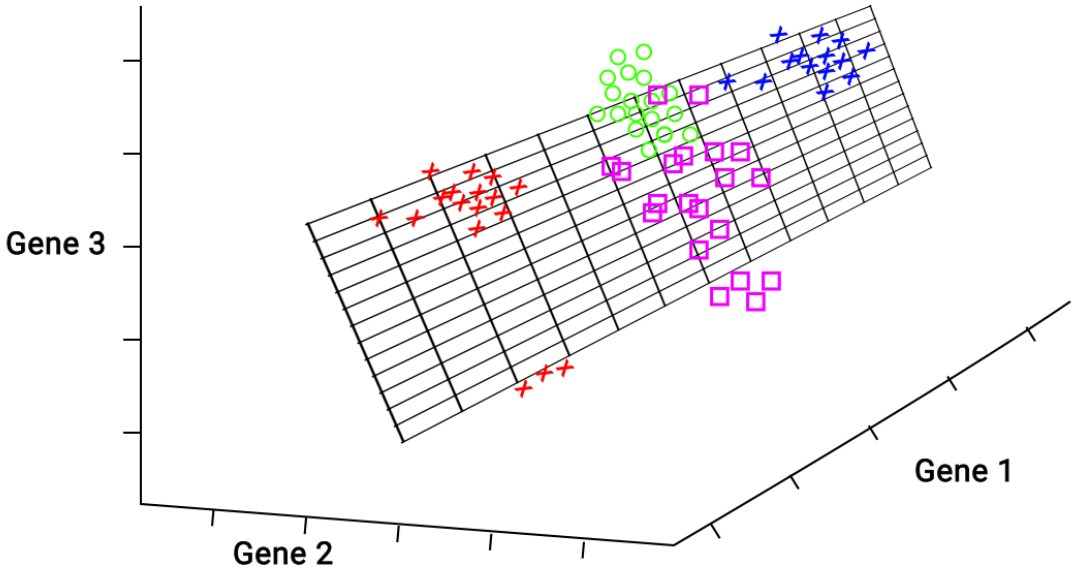
The 2-Dimensional Case



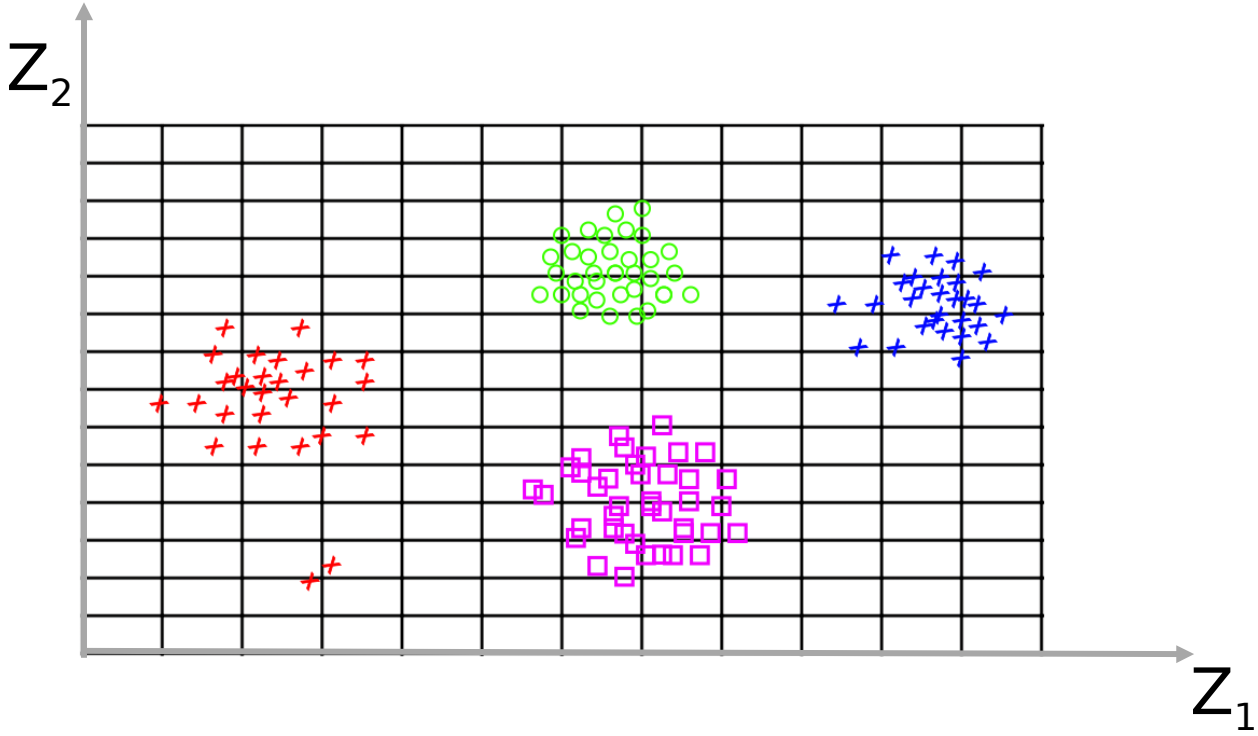
- Find Z_1 and Z_2
- Take projection of each data point along
- Z_1 is called the **Principal Component**

The 3-Dimensional Case

The 3-Dimensional Case



The 3-Dimensional Case



The n-Dimensional Case

The n-Dimensional Case

- n features f_1, f_2, …, f_n → k directions Z_1, Z_2, …, Z_k
- $k < n$
- k Principal Components
- Meaning of new features is not interpretable

Thank
You!

Analytics Vidhya
Learn everything about analytics