This module covers…

- Plot Diagrams of low dimensional data sets like Box Plot, Run Chart, Scatter Plot and Histograms
- Draw conclusions out of the diagrams you've plotted
- Reduce dimensions of your data set
- Understand the concept of Phase Diagrams

In this Video you will learn…

Plot Diagrams of low dimensional data sets like Box Plot, Run Chart, Scatter Plot and Histograms

matplotlib

- plotting library for python
- open source
- de-facto standard for plotting in python

Sampling

- what to do if you have more data than memory
- plotting library expects vectors and matrices
- should fit into memory
- not made for big data
- sampling
  - preserves properties of data but
  - reduces computational cost

box plots

- show many properties at the same time
  - mean
  - standard deviation
  - skew
  - outliers (kurtosis)
- distribution of your data (vector)

run charts

- the ones you know from stock market
- perfectly suited for time series
- x-axis always is time
- y-axis observed value over time

scatter plots

- individual data points, addressed by
  - two or thee dimensions
- classification boundaries
- clusters
- anomalies

histograms

- get an idea of distribution of values
- find regions of high and low value concentration

Summary

- low dimensional data can be plotted using charts
- depending on the chart types used, these properties can be visually observed
  - mean, variance, skew
  - clustering, outliers and means of separation
- although each plot has a clear purpose…
- …often randomly playing with plots does the job best

The next module covers…

Dimensionality Reduction