Time Series

Time Series is a sequence of real values measuring some variable **over time**, usually at **uniform intervals**.

In **Multivariate Time Series** the goal is to study **multiple variables** that change over time.

Time Series Profiling

- **Dimensionality** - number of elements in the sequence;
- **Granularity** - time interval between each element;
- **Components** - time series can be decomposed in its components:
  - **Trend** - general variation over time;
  - **Seasonal** - periodic variation over time - $>=$ 1 year;
  - **Cyclical** - periodic variation over time - $<$ 1 year;
  - **Residual/Noise** - irregularities and fluctuation that occur along time, are random, have a short length and are non-repetitive;
- **Stationarity** - a time series is stationary if its statistical properties (mean, variance, covariance) do not change over time;
  - There is **no trend or seasonality**;
  - The easiest way to check for stationarity is to plot the data and check for trends or seasonality;
  - The **Augmented Dickey-Fuller (ADF) test** is a statistical test for stationarity;
    - $p-value \leq 0.05$ - the time series is **stationary**;
    - $p-value > 0.05$ - the time series is **not stationary**;

Distance Measures between Time Series

- We need to compare the values observed in every instant of time, which can be done by viewing each series as a point in R^n, where n is the number of observations;
  - Then, we can use the **Euclidean distance** to compare the series;

Some of these **time distortions** are particularly relevant:

- **Vertical Offset Translation** - series present similar behavior, but with different vertical offsets;
- **Horizontal Offset Translation-** series present similar behavior, but with different time lags;
- **Amplitude Scaling** - series present similar behavior, but with different amplitudes;
- **Linear Trend** - series present similar behavior, but with different linear trends;
- **Noise** - series present similar behavior, but with different noise levels.

Dynamic Time Warping (DTW)

- Distance measure that allows for **time distortions**;
- Recursively computes the **optimal alignment** between two time series;
- Uses Euclidean distance as a local distance measure;
- $DTW_{\text{distance}} = \frac{1}{N} min(\sqrt{\sum_{n=1}^{N}(x_i - y_i)^2})$;
- Construct a matrix M of size $N \times M$, where N and M are the number of points in the two time series;
  - $M_{i,j} = (x_i - y_j)^2 + min(M_{i-1,j}, M_{i,j-1}, M_{i-1,j-1})$;
  - $DTW_{\text{distance}} = \sqrt{M_{N,M}}$.

Time Series Transformation

- **Sliding Window** - start at the beginning of the time series, apply a function to the first k points, then move the window one step forward and repeat the process, until there are no more points in the time series;
- **Moving Average** - average of the values in the window;

  - Used to **smooth** the time series;
  - Applying **average in sliding window approach**;
  - **Exponential Moving Average** - weighted average of the values in the window, where the weights decrease exponentially as we move back in time;

- **Normalization** - addresses the **amplitude scaling problem**; the goal is to scale the values of the time series to a fixed range, usually $[0, 1]$ or $[-1, 1]$;
- **Trend Removal** - Compute the straight line that best fits the original time series and then subtract the former from the latter (the result is a time series with no trend);
- **Differencing** - Compute the difference between consecutive values of the time series (the result is a time series with no trend and no seasonality);
- **Segmentation** - Split the time series into segments and then apply a function to each segment;
  - **Piecewise Aggregate Approximation (PAA)** - split the time series into k segments and then compute the average of each segment; **fast, simple and intuitive**;
  - **Symbolic Aggregate Approximation (SAX)** - split the time series into k segments and then map each segment to a symbol, based on the average of the segment; **linear time to compute**; fewer bits to encode symbols;
  - **Singular Value Decomposition (SVD)** - represent a time series as a **linear composition of eigen waves**;
  - **Discrete Fourier Transform (DFT)** - represent a time series as a **linear composition of sine and cosine waves**;
  - **Discrete Wavelet Transform (DWT)** - represent a time series as a **linear composition of wavelets** - functions that cut the sequence into different frequency components.