TIME SERIES FORECASTING

FORECASTING → Prevision

Given a time series

$x_i, x_{i-1}, \dots$

We want to estimate the future values in the series, i.e. we want

$$\hat{x}_{i+D} = f(\underbrace{x_i, x_{i-1}, \dots, x_{i-w+1}}_{\text{historical memory used to estimate new value.}})$$

$D \geq 1$

$w :=$ Window size

There is a connection between TIME SERIES FORECASTING and REGRESSION.

REGRESSION uses pairs of measurements $(x_1, y_1), \dots, (x_N, y_N)$ to build a model $f()$ such that

$$\underset{\text{RESPONSE}}{y_i} = f(x_i)$$

There are four option, since x and y can both be either SCALARS or VECTORS

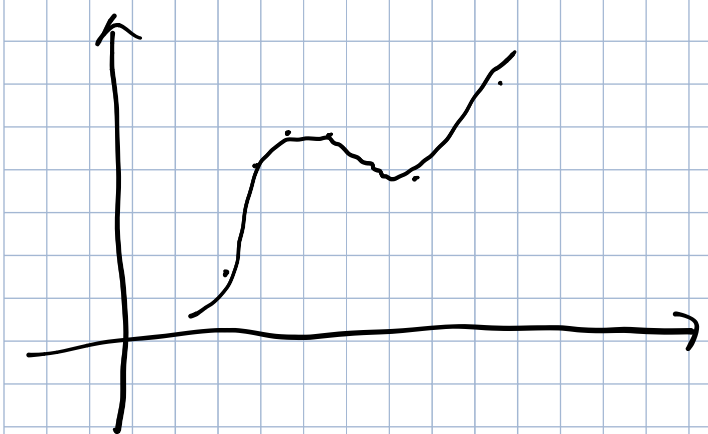
X → SCALAR, SIMPLE REGRESSION
X → VECTOR, MULTIPLE REGRESSION
Y → SCALAR, UNIVARIATE REGRESSION
Y → VECTOR, MULTIVARIATE REGRESSION

OSS: Classification can be formalized as a

If we have multiple classes for each row of our dataset then we can have MULTIPLE - MULTIVARIATE REGRESSION.

Time series forecasting problem can be formalized as a MULTIPLE - UNIVARIATE regression problem.

Regression problem can be thought of as "curve fitting" problem, since we wanto to find the curve that intercepts our points.



If we see the x-AXIS as time, then we can think about REGRESSION as a method that let's us extrapolate info. from the past.

x = time index.

Consider a random process

$X_1, \dots X_i, X_{i+1}, \dots$

if we fix the time at i
we have a set of r.v. $\{X_{i-1}, \dots, X_1\}$

!

Model to do TIME SERIES FORECASTING

$$\hat{X}_{i+D} = f(X_i, X_{i-1}, \dots)$$

A good estimator is one that maximizes the expected value

$$\mathbb{E}[\hat{X}_{i+D} \mid X_i, X_{i-1}, \dots]$$

If we work with observations, then we can define the POINT-WISE ERROR as

$$e_i := \hat{x}_i - x_i$$

By computing the point-wise error at different points during our forecasting then we can define,

- MEAN SQUARED ERROR,

$$MSE = \frac{1}{N} \cdot \sum_{j=1}^{N} e_i^2$$

MOST COMMONLY USED!

- MEAN ABSOLUTE DIFFERENCE

$$MAD = \frac{1}{N} \sum_{j=1}^{N} |e_i|$$

COMMON PREDICTION METHODS

$$\hat{x}_{i+D} = f(x_i, x_{i-1}, \dots)$$

- NAÏVE - 1

$$\hat{x}_{i+D} := x_i$$

Good if there is no trend, or if D is small.

- **NAÏVE - 2**

$$\hat{x}_{i+D} := x_i \left(1 + \frac{x_i - x_{i-1}}{x_{i-1}} \right)$$

Not good if there is noise

- **MOVING AVERAGE**

$$\hat{x}_{i+D} := \frac{1}{w} \sum_{j=i-w+1}^{i} x_i$$

Good for smoothing the noise, but it does not take into account the trend.

All w previous points have all the same importance in the basic version of MOVING AVERAGE.

- Weighted Moving Average

$$\hat{x}_{i+D} := \frac{1}{W} \sum_{j=i-w+1}^{i} x_j \cdot w_j$$

A weight w_j is introduce to scale the importance of each observation.

Generally,

Further back in time $\rightarrow$ less importance

OSS: Consider the following LINEAR TIME-INVARIANT ? FILTER, which is a linear convolution

$$y(i) = \sum_{k=0}^{N} h(k) \cdot (i-k)$$

(LPF)
LINEAR PREDICTION FILTER

$$X \rightarrow \boxed{H} \xrightarrow{Y}$$

IMPULSE RESPONSE OF THE FILTER

We can map WMA as one of such filters.

SINGLE EXPONENTIAL SMOOTHING

In this case we have,

$$\hat{x}_{i+D} = \alpha \cdot x_i + \alpha(1-\alpha)\, x_{i-1} + \alpha(1-\alpha)^2\, x_{i-2} + \dots$$

$$\hat{x}_{i+D} = \alpha\, x_i + (1-\alpha) \cdot \hat{x}_i$$

where $\alpha \in (0, 1)$ is a SMOOTHING PARAMETER.

Not that good for time series with a very steep trend.

Remove this?

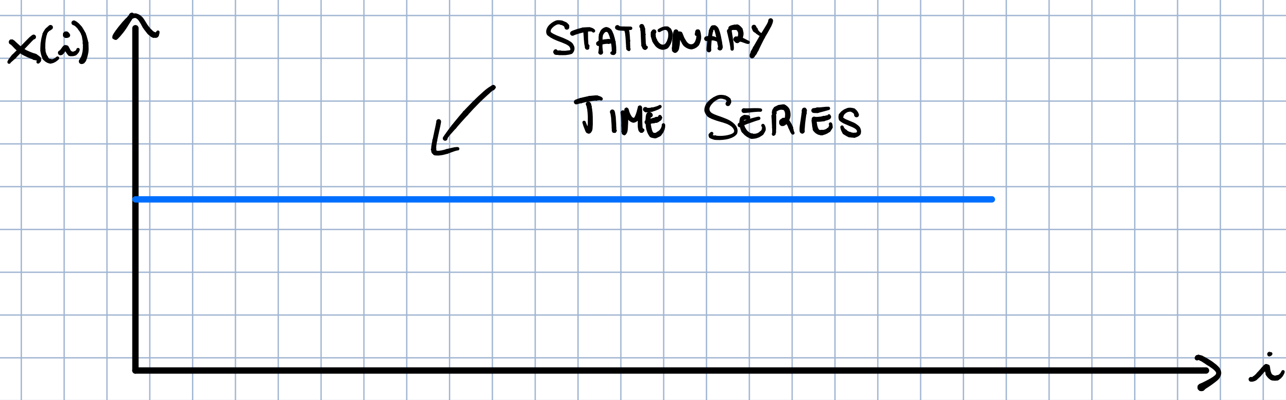
Q: time series "with a trend" are stationary or not?

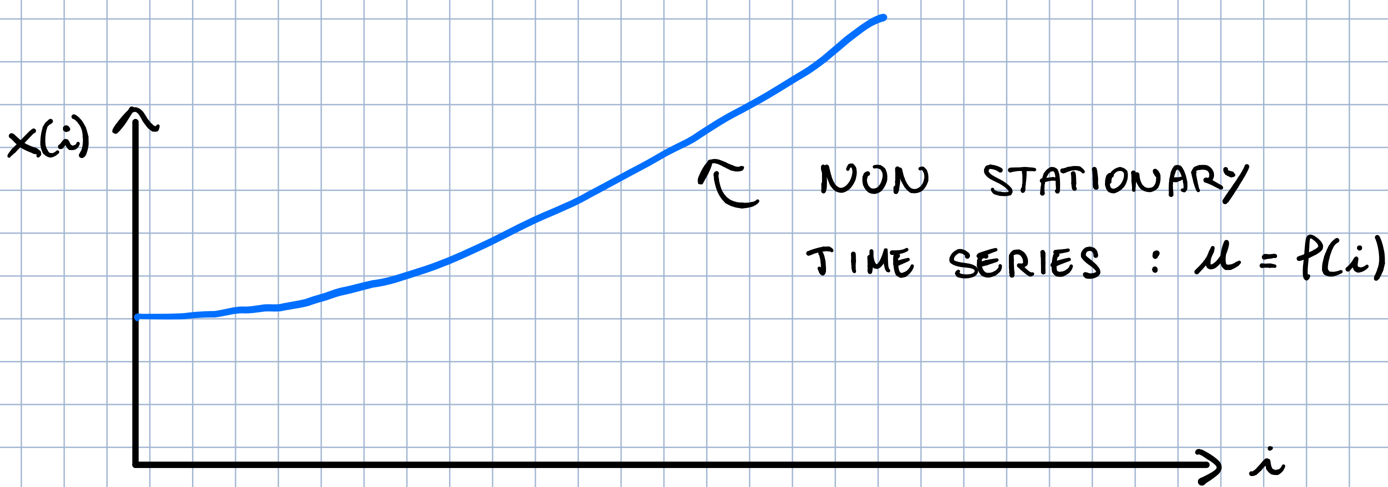
R: A time series is said to be STATIONARY in the wide sense if

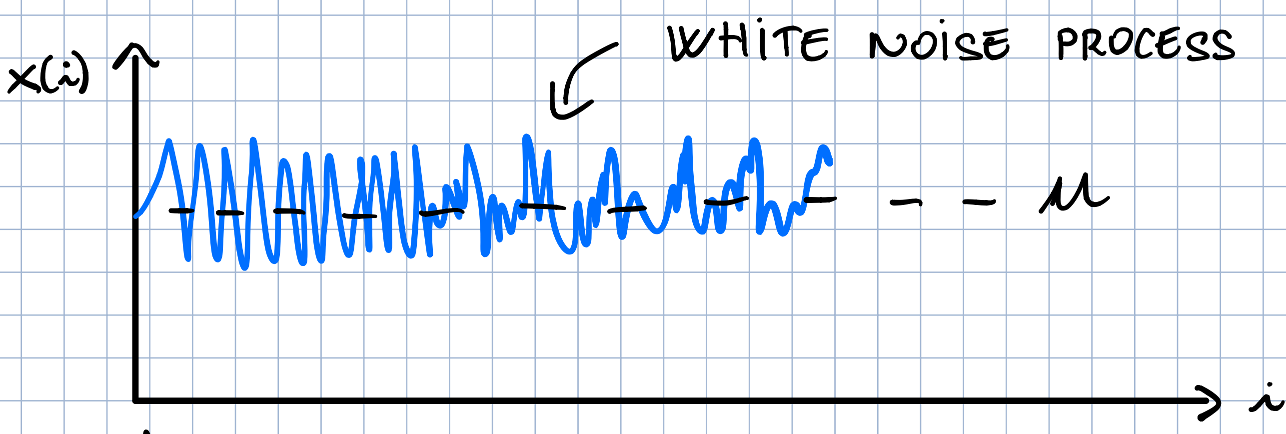
i) μ = constant

ii) $C = f(\tau)$, $\tau = t_1 - t_2$

Auto-Covariance for a given length of time



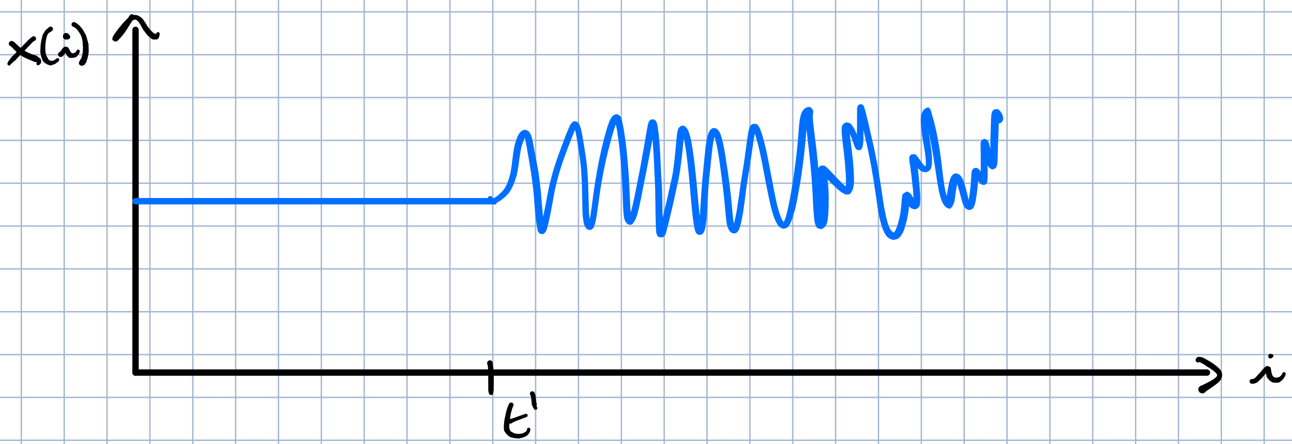




Stationary time series since:

1) The mean seems constant

2) The similarity does not seem to change with time



This time series is NOT STATIONARY since before t' the autocovariance for a given τ is high, while after t' is low. Therefore the autocovariance is not just a function of τ, but also of time.

The idea is then to study the time series and remove the trend component from it. After that we may estimate the seasonal component and remove it

(!)

OSS : We cannot predict NOISE, because the auto-correlation decreases fast with the increase of τ.

The objective of time series analysis is to build a model which has a MSE similar to the variance of the irregular component.

ASSIGNMENT n° 6

Write a report discussing each of the previous assignments