time series analysis: Uncover patterns in historical data that change over time
The goals of time series analysis:
- Identify and model the structure of the time series.
- Forecast future values in the time series.

Multicollinearity happens when two or more features (or variables) in a dataset are very similar or highly related to each other. In simple terms, it means that some features provide the same information, making it harder for a model to understand which one is actually important.

Lecture-9

Components of Time Series:

1. Trend: up trend
   down trend
   horizontal trend

Trend: A trend is nothing but a movement to relatively higher or lower values over a long period.

2. Seasonality — It's a repeating pattern within a fixed period, jatka fish
3. Irregularity — Irregularity typically occurs for a brief period and does not repeat, covid
4. Cyclic — repeating up and down movements, so this means we can go over more than a year. Cyclic does not have any fixed patterns. They can happen anytime, like in a year in a decade, or maybe within six months.

- ARIMA (Autoregressive Integrated Moving Average)

The ARIMA model is a powerful statistical method used to analyze and forecast time series data by combining autoregression, differencing and moving average models. It is particularly useful when the data exibits patterns such as trends, cycles, seasonality, which need to be accounted for before making prediction.

A time series is stationary if it's statistical properties such as mean, variance and autocovarice are constant over time.

AR model:

The autoregressive (AR) model is a type of time series model that predicts future values based on past values of the same series.

For example: in an AR(1) model, the value of y_t is related to the immediate past value y_{t-1}:

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} \pm \phi_p y_{t-p} + \epsilon_t \quad \text{---} (1)$$

AR(P)

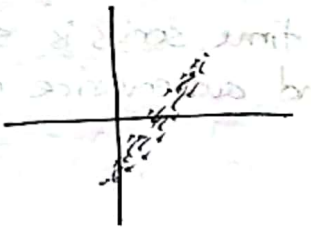
y_t = is the value of the time series at time t

δ = is a constant term added when the time series is not centered around zero (i.e. the mean is non-zero)

ϕ = are the AR coefficients that determines the influence of past values $Y_{t-1}, Y_{t-2} \cdots Y_{t-p}$ on the current time y_t

ϵ_t = Random error term (White noise)

Equation (1) is a non-zero centered series. To convert it into zero centered, we need to remove the constant term (mean). δ or mean is substracted from each data points, it helps making the series detrened and zero centered



Non Zero Centered

The equation becomes;

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_p y_{t-p} + \epsilon_t$$

Zero Centered

Why Zero-Centering is useful:

① Simplifies the AR model: By removing δ, the equation becomes easier to interpret, focusing on previous and current values.

② Stationary: Zero-centering is a step toward achieving stationarity by eliminating trends or constant shifts of data.

Example:

The original time series data y_t is:

$y_1 = 12$, $y_2 = 14$, $y_3 = 16$, $y_4 = 13$, $y_5 = 15$

$\Rightarrow$ mean, $\bar{y} = \frac{12+14+16+13+15}{5} = 14$

Converting into zero center;

$y_1' = 12 - 14 = -2$, $y_2' = 14 - 14 = 0$, $y_3' = 16 - 14 = 2$, $y_4' = 13 - 14 = -1$, $y_5' = 15 - 14 = 1$

Assume for AR(2) model;

$$y_t = 0.5\, y_{t-1} + 0.3\, y_{t-2} + \epsilon_t$$

For Prediction of y_6 $= 0.5 \times y_5 + 0.3 \times y_4 + \epsilon_t$

$= 0.5 \times (1) + 0.3 \times (-1) + 0.2$

$= 1.0$

MA Model: Moving Average is a type of time series model that uses past forecast errors rather than just past values of the time series itself to predict future values. It assumes that the value at time t depends on the random error at the time t, as well as previous errors. MA model is already zero centered.

$$y_t = \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_q \epsilon_{t-q}$$

MA(q)

AP(p)

Example: (MA(1)) = MA(2)

For $y_t = \epsilon_t + \phi_1 \epsilon_{t-1}$

Here, $\phi_1 = -0.5$

$\epsilon_t = 4$

Day	Prediction	Actual Value	Error
1	–	4	–
2	4	10	4-10 = -6
3	7	14	7-14 = -7
4	7.5	9	7.5-9 = -1.5

$y_t = 4 + (-6 \times -0.5) = 4+3 = 7$

$y_t = 4 + (-7 \times -0.5) = 7.5$

- **Autocorrelation Function (ACF)**: measures the total correlation between a time series and its past values.

The ACF shows the correlation between Y_t and its lagged (previous) values $y_{t-1}, y_{t-2}, \ldots$.

The ACF captures both direct and indirect correlation. As we already know MA model requires the previous direct and indirect effects of the past values for the prediction of ~~present~~ future value, ACF provides the exact measure for it. So, ACF is used by MA model.

- **Partial Autocorrelation Function (PACF)**: measures only the direct correlation. PACF Lag 1 will show a strong positive correlation because y_t is directly related to y_{t-1}. But PACF Lag 2 will show zero correlation. The PACF helps isolate the direct relationship between the current value and its past values by removing the influence of intermediate lags. That's why ~~PCA~~PACF is used by AR model.

In the context of time series analysis, **PACF (Partial Autocorrelation Function) Lag 1 = 1** means:

- The **first lag** of the time series is **highly correlated** with the current value.
- In other words, the most recent past value has a **strong direct influence** on the current value, even after removing the effects of other intermediate lags.

This suggests that the time series likely has an **AR(1)** (Auto-Regressive model of order 1) component, where only the immediate past value matters significantly for prediction.

- **ARMA**: Combination of AR and MA model.

$$y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_1 + \theta_1 \epsilon_{t-1} + \ldots \theta_q \epsilon_{t-q}$$

Quiz-3 (Solve)

Set-A

1

From Figure-1;

AR uses PACF

only one lag outside cutoff.

$\therefore AR(1) = AR(P)$

From Figure-2;

MA uses ACF

only one lag outside cutoff.

$\therefore MA(1) = MA(q)$

Set-B

From Figure-1;

AR uses PACF

3 lag outside cutoff.

$\therefore AR(3) = AR(P)$

From Figure-2;

MA uses PACF

2 lag outside cutoff

$\therefore MA(2) = MA(q)$

Set-D

From figure-1; ACF chart shows several significant spikes in ACF values. The ACF values starts to decreases at 12, 24, 36 and 48. This pattern represents Autoregressive Pattern of every 12 months.

From the PACF, lag 12 is quite large, but values of 24, 36, 48 are close to zero. So, AR(1) with period = 12 will be considered.

From figure-2; ACF chart
12, 24, 36 and 48 has been addressed by the seasonal AR(1) term. Only significant value is at lag 1. ACF cuts off sharply at lag 1. On the other hand, PACF plot in figure exhibits a slowly decaying PACF. MA(1) model should be considered for the nonseasonal portion of SARIMA.

$$\underbrace{(0,1,1)}_{\text{non-seasonal}} \times \underbrace{(1,0,0)_{12}}_{\text{Seasonal}}$$

Set-C

The exponential decay in ACF suggests AR component in the model. The PACF shows a significant spike at lag 1 and then cuts off, suggests AR(1) model.

S=4; The regular spikes every 4 lags in both ACF and PACF indicates data has seasonal pattern with a periodicity of 4.

The PACF shows significant spikes at lags 4, 8, 12 which are spread at regular intervals, these represent the presence of a Seasonal AR component. Since 1 spike after every 4 interval. So, Seasonal AR(1).

As the series has seasonality, we apply seasonal differencing (D=1).

In ACF, we can see significant spikes at 4, 8, 12 and 16 which indicates, seasonal moving average, suggests Seasonal MA(1).

$$(1,0,0) \times (1,1,1)_4$$