Go Back to Practical Data Science

Course Content

Auto Correlation and Partial Auto Correlation Functions

In the previous pre-read, we learned about the ARIMA model. The ARIMA model requires 3 necessary parameters to model the time series data. The **d** value can be determined from the number of times the data has had past values subtracted. To get the **p** and **q** values, however, we use the ACF and PACF plots.

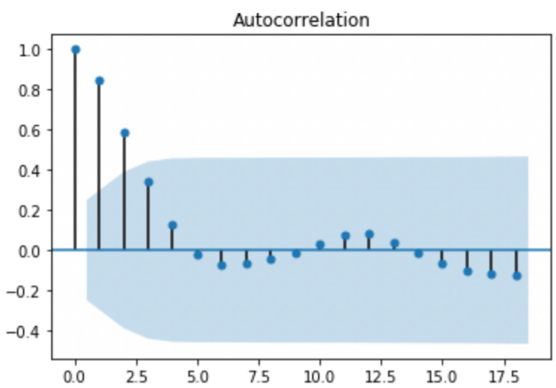
The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) plots of the series are necessary to determine the order of AR(p) and MA(q) terms. Although ACF and PACF do not directly dictate the order of the ARIMA model, the plots can facilitate understanding of the order and provide an idea of which model may be a good fit for the Time Series data.

Auto Correlation Function (ACF)

The bar chart of coefficients of correlation between a time series and its lagged values is known as the ACF plot. In simple terms, ACF explains how the present value of a given time series is correlated with its past values (1-unit past,...., n-unit past).

Assume, $y(t-1)$,...., $y(t-n)$ are values of a time series at the time t, $t-1$,...,$(t-n)$, respectively, then the correlation coefficient between $y(t)$ and $y(t-1)$ is **lag 1**, the correlation coefficient between $y(t)$ and $y(t-2)$ is **lag 2,** and so on.

An example of an ACF plot is observed below:



In this plot, the y-axis represents the correlation coefficient, while the x-axis represents the number of lags. The blue area represents the confidence interval indicating the statistical significance of the correlation.

Estimation of q in MA model:

- We can estimate the MA(q) using the ACF plot
- It can be estimated by observing how many lags are above or below the confidence interval before the next lag enters the blue area

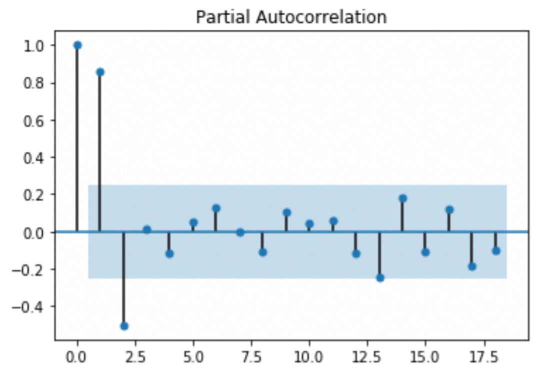
From the above figure, we can estimate the MA term as 2 because of the highest lag at which the plot extends beyond the confidence interval. After the value at lag 2, every value is relatively small and within the confidence interval, indicating the lack of a statistically significant correlation. So, the data can be modeled well with MA(2).

Partial Auto Correlation Function (PACF)

PACF is a partial autocorrelation function that describes the partial correlation between the series and its own lag. In other words, the partial correlation for each lag is the unique correlation between those two observations after taking out the intervening correlations.

For example, the partial autocorrelation for lag 3 is only the correlation that lags 1 and 2 do not explain. In mathematical terms, we correlate the parts of $y(t)$ and $y(t-3)$ that are not predicted by $y(t-1)$ and $y(t-2)$.

An example of a PACF plot is observed below:



Estimation of p in AR model:

PACF plots are used to estimate the value of the AR term. First, ignore the value at lag 0. It will always show a perfect correlation since we are estimating the correlation between the present value and itself. The same logic as the MA plot applies here, how many lags are above or below the confidence interval before the next lag enters the blue area.

In the above figure, we can estimate the AR term as 2, because of the highest lag at which the plot extends beyond the confidence interval. After the value at lag 2, every value is relatively small and within the confidence interval. So, the data can be modeled well with AR(2).

Happy Learning!

Previous

Next

Help