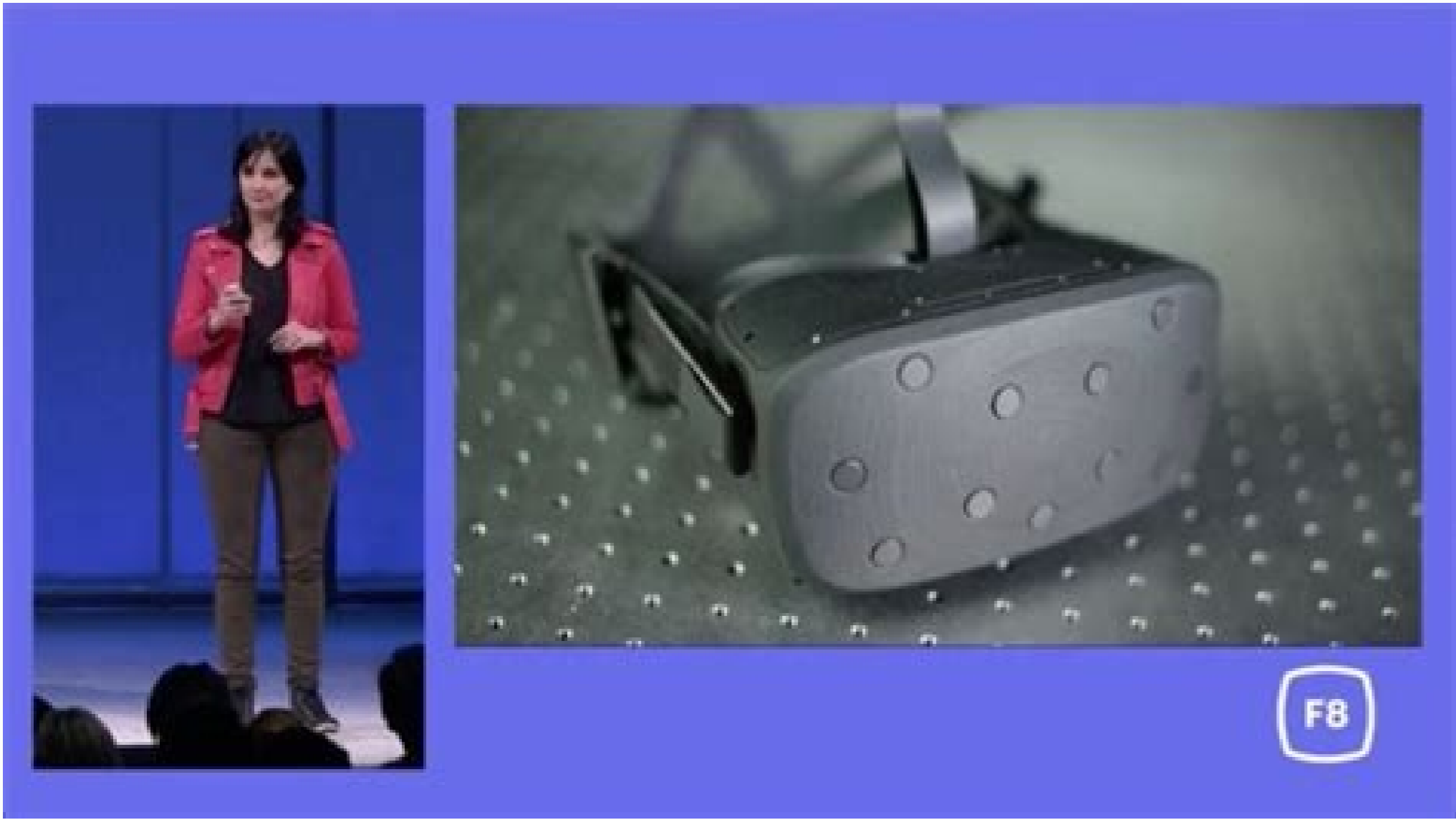


I'm not robot!



TOPICAL PAST PAPERS
MAY 2012 - OCT 2020

SPONSORED BY
Education House International Academy
www.educationhouse.com.my

exammate
www.exammate.com

IGCSE Cambridge
ICT

0417 PAPER 1
QUESTIONS + ANSWERS

Types of time series analysis pdf. Different types of time series analysis. Types of time series analysis methods. 4 types of variation in time series analysis. Different types of error in time series analysis. Types of time series analysis in statistics. Four types of variation that is estimated in time-series analysis. Three types of time series analysis.

This article was published as a part of the Data Science Blogathon. This article was published as a part of the Data Science Blogathon. A Time-Series represents a series of time-based orders. It would be Years, Months, Weeks, Days, Hours, Minutes, and Seconds. A time series is an observation from the sequence of discrete-time of successive intervals. A time series is a running chart. The time variable/feature is the independent variable and supports the target variable to predict the results. Time Series Analysis (TSA) is used in different fields for time-based predictions - like Weather Forecasting, Financial, Signal processing, Engineering domain - Control Systems, Communications Systems. Since TSA involves producing the set of information in a particular sequence, it makes a distinct from spatial and other analyses. Using AR, MA, ARMA, and ARIMA models, we could predict the future. Introduction to Time Series Analysis Time Series Analysis is the way of studying the characteristics of the response variable with respect to time, as the independent variable. To estimate the target variable in the name of predicting or forecasting, use the time variable as the point of reference. In this article we will discuss in detail TSA Objectives, Assumptions, Components (stationary, and Non-stationary). Along with the TSA algorithm and specific use cases in Python. What is Time Series Analysis (TSA) and its assumption? How to analyze? Time Series Analysis Significance and its types. Components of Time Series What are the limitations of time series? Detailed Study of Time Series Data types. Discussion on stationary and Non-stationary components Conversion of Non-stationary into stationary Why is Time Series Analysis used in Data Science and Machine Learning? Time Series Analysis in Data Science and Machine Learning Implementation of Moving Average (WEIGHTS - SIMPLE MOVING AVERAGE) Understanding ARMA and ARIMA Implementation steps for ARIMA Time Series Analysis - Process flow (Re-gap) What is Time Series Analysis Definition. If you see, there are many more definitions for TSA. But make it simple. A time series is nothing but a sequence of various data points that occurred in a successive order for a given period of time Objectives: To understand how time series works, what factors are affecting a certain variable(s) at different points of time. Time series analysis will provide the consequences and insights of features of the given dataset that changes over time. Supporting to derive the predicting the future values of the time series variable. Assumptions: There is one and the only assumption that is "stationary", which means that the origin of time, does not affect the properties of the process under the statistical factor. Quick steps here for your reference, anyway. Will see this in detail in this article later. Collecting the data and cleaning it Preparing Visualization with respect to time vs key feature Observing the stationarity of the series Developing charts to understand its nature. Model building - AR, MA, ARMA and ARIMA Extracting insights from prediction TSA is the backbone for prediction and forecasting analysis, specific to the time-based problem statements. Analyzing the historical dataset and its patterns Understanding and matching the current situation with patterns derived from the previous stage. Understanding the factor or factors influencing certain variable(s) in different periods. With help of "Time Series" we can prepare numerous time-based analyses and results. Forecasting Segmentation Classification Descriptive analysis Intervention analysis Trend Seasonality Cyclical Irregularity Trend: In which there is no fixed interval and any divergence within the given dataset is a continuous timeline. The trend would be Negative or Positive or Null Trend Seasonality: In which regular or fixed interval shifts within the dataset in a continuous timeline. Would be bell curve or saw tooth Cyclical: In which there is no fixed interval, uncertainty in movement and its pattern Irregularity: Unexpected situations/events/scenarios and spikes in a short time span. What are the limitations of

Time Series Analysis? Time series are the below-mentioned limitations, we have to take care of those under our analysis, Similar to other values, not supported by TSA in the data points must be linear in their relationship. Data transformations are mandatory, so a little expensive. Most series are in the form of time series. The MEAN value of them should be completely constant in the data during the analysis. The VARIANCE should be constant with the time-frame. The COVARIANCE measures the relationship between two variables. 6.2 Non- Stationary; This is just the opposite of Stationary. During the TSA model preparation workflow, we must access if the given dataset is Stationary or NOT. Using Statistical and Plots test. 7.1 Statistical Test: There are two tests available to test if the dataset is Stationary or NOT. Augmented Dickey-Fuller (ADF) Test Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test 7.1.1 Augmented Dickey-Fuller (ADF) Test or Unit Root Test: The ADF test is the most popular statistical test and with the following assumptions. Null Hypothesis (H0): Series is non-stationary Alternate Hypothesis (HA): Series is stationary p-value >0.05 Fail to reject (H0) p-value Smoothing Factor. It has a value between 0,1. Represents the weighting applied to the very recent period. Lets will apply the exponential moving averages with a smoothing factor of 0.1 and 0.3 in the given dataset. # EMA Air Temperature # Let's smoothing factor - 0.1 df temperature'EMA_0.1'] = df_temperature.average_temperature.ewm(alpha=0.1, adjust=False).mean() # Let's smoothing factor - 0.3 df_temperature['EMA_0.3'] = df_temperature.average_temperature.ewm(alpha=0.3, adjust=False).mean() # green - Avg Air Temp, red- smoothing factor - 0.1, yellow - smoothing factor - 0.3 colors = ['green', 'red', 'yellow'] df_temperature['average_temperature', 'EMA_0.1', 'EMA_0.3']].plot(color=colors, linewidth=3, figsize=(12,6), alpha=0.8) plt.xticks(fontsize=14) plt.yticks(fontsize=14) plt.legend(labels=['Average air temperature', 'EMA - alpha=0.1', 'EMA - alpha=0.3'], fontsize=14) plt.title('The yearly average air temperature in city', fontsize=20) plt.xlabel('Year', fontsize=16) plt.ylabel('Temperature [°C]', fontsize=16) When dealing with TSA in Data Science and Machine Learning, there are multiple model options are available. In which the Autoregressive-Moving-Average (ARMA) models with [p, d, and q]. P==> autoregressive lags q== moving average lags d==> difference in the order Before we get to know about Arima, first you should understand the below terms better. Auto-Correlation Function (ACF) Partial Auto-Correlation Function (PACF) 10.1 Auto-Correlation Function (ACF). ACF is used to indicate and how similar a value is within a given time series and the previous value. (OR) It measures the degree of the similarity between a given time series and the lagged version of that time series at different intervals that we observed. Python Statsmodels library calculates autocorrelation. This is used to identify a set of trends in the given dataset and the influence of former observed values on the currently observed values. 10.2 Partial Auto-Correlation (PACF); PACF is similar to Auto-Correlation Function and is a little challenging to understand. It always shows the correlation of the sequence with itself with some number of time units per sequence order in which only the direct effect has been shown, and all other intermediary effects are removed from the given time series. Auto-correlation and Partial Auto-Correlation plot acf(df_temperature) plt.show() plot acf(df_temperature, lags=30) plt.show() Observation: The previous temperature influences the current temperature, but the significance of that influence decreases and slightly increases from the above visualization along with the temperature with regular time intervals. 10.3 Types of Auto-correlation 10.4 Interpret ACF and PACF plots ACF PACF Perfect ML -Model Plot declines gradually Moving Average model Plot drops instantly Plot declines gradually Plot decline gradually Plot Decline gradually ARMA Plot drop instantly Plot drop instantly You wouldn't perform any model Remember that both ACF and PACF require stationary time series for analysis. Now, we learn about the Auto-Regressive model This is a simple model, that predicts future performance based on past performance. mainly used for forecasting, when there is some correlation between values in a given time series and the values that precede and succeed (back and forth). An AR model is a Linear Regression model, that uses lagged variables as input. The Linear Regression model can be easily built using the scikit-learn library by indicating the input to use. Statsmodels library is used to provide autoregression model-specific functions where you have to specify an appropriate lag value and train the model. It is provided in the AutoReg class to get the results. using simple steps Creating the model AutoReg() Call fit() to train it on our dataset. Returns an AutoRegResults object. Once fit, make a prediction by calling the predict () function The equation for the AR model (Let's compare Y=mX+c) Yt =C+b1 Yt-1 + b2 Yt-2+.....+ bp Yt-p+ Ert Key Parameters p=past values Yt=Function of different past values Ert=errors in time C=intercept Lets's check, given data-set or time series is random or not from matplotlib import pyplot from pandas.plotting import lag_plot lag_plot(df_temperature) pyplot.show() Observation: Yes, looks random and scattered. Implementation of Auto-Regressive model #import libraries from matplotlib import pyplot from statsmodels.tsa.ar_model import AutoReg from sklearn.metrics import mean_squared_error from math import sqrt # load csv as dataset #series = read_csv('daily-min-temperatures.csv', header=0, index_col=0, parse_dates=True, squeeze=True) # split dataset for test and training X = df_temperature.values train, test = X[1:len(X)-7], X[len(X)-7:] # train autoregression model = AutoReg(train, lags=20) model fit = model.fit() print('Coefficients: %s' % model.fit.params) # Predictions predictions = model.fit.predict(start=len(train), end=len(train)+len(test)-1, dynamic=False) for i in range(len(predictions)): print('predicted=%f, expected=%f' % (predictions[i], test[i])) rmse = sqrt(mean_squared_error(test, predictions)) print('Test RMSE: %.3f' % rmse) # plot results pyplot.plot(test) pyplot.plot(predictions, color='red') pyplot.show() OUTPUT predicted=15.893972, expected=16.275000 predicted=15.917959, expected=16.600000 predicted=15.812741, expected=16.475000 predicted=15.787555, expected=16.375000 predicted=16.023780, expected=15.940271, expected=16.525000 predicted=15.831538, expected=16.758333 Test RMSE: 0.617 Observation: Expected (blue) Against Predicted (red). The forecast looks good on the 4th and the deviation on the 6th day. Implementation of Moving Average (WEIGHTS - SIMPLE MOVING AVERAGE) import numpy as np alpha= 0.3 n = 10 w_sma = np.repeat(1/n, n) colors = ['green', 'yellow'] # weights - exponential moving average alpha=0.3 adjust=False w_ema = [(1-ALPHA)**i if i==N-1 else alpha*(1-alpha)**i for i in range(n)] pd.DataFrame({'w_sma': w_sma, 'w_ema': w_ema}).plot(color=colors, kind='bar', figsize=(8,5)) plt.xticks(1) plt.yticks(fontsize=10) plt.legend(labels=['Simple moving average', 'Exponential moving average (alpha=0.3)'], fontsize=10) # title and labels plt.title('Moving Average Weights', fontsize=10) plt.ylabel('Weights', fontsize=10) Understanding ARMA and ARIMA ARIMA This is a combination of the Auto-Regressive and Moving Average model for forecasting. This model provides a weakly stationary stochastic process in terms of two polynomials, one for the Auto-Regressive and the second for the Moving Average. ARMA is best for predicting stationary series. So ARIMA came in since it supports stationary as well as non-stationary. AR ==> Uses the past values to predict the future MA ==> Uses the past eror terms in the given series to predict the future I==> uses the differencing of observation and makes the stationary data AR+I+MA= ARIMA Understand the Signature of ARIMA p==> log order ==> No of lag observations. d==> degree of differencing ==> No of times that the raw observations are differenced. q==>order of moving average ==> the size of the moving average window Implementation steps for ARIMA Step 1: Plot a time series format Step 2: Difference to make stationary on mean by removing the trend Step 3: Make stationary by applying log transform. Step 4: Difference log transform to make as stationary on both statistic mean and variance Step 5: Plot ACF & PACF, and identify the potential AR and MA model Step 6: Discovery of best fit ARIMA model Step 7: Forecast/Predict the value, using the best fit ARIMA model Step 8: Plot ACF & PACF for residuals of the ARIMA model, and ensure no more information is left. Implementation of ARIMA Already we have discussed steps 1-5, let's focus on the rest here. from statsmodels.tsa.arima_model import ARIMA model = ARIMA(df_temperature, order=(0, 1, 1)) results_ARIMA = model.fit() results_ARIMA.summary() results_ARIMA.forecast(3)[0] Output array([16.47648941, 16.48621826, 16.49594711]) results_ARIMA.plot_predict(start=200) plt.show() Recurrent Neural Networks is the most traditional and accepted architecture, fitment for Time-Series forecasting based problems. RNN is organized into successive layers and divided into Each layer has equal weight and every neuron has to be assigned to fixed time steps. And remember that every one of them is fully connected with a hidden layer (input and Output) with the same time steps and the hidden layers are forwarded and time-dependent in direction. Components of RNN Input: The function vector of x(t) is the input, at time step t. Hidden: The function vector h(t) is the hidden-state at time t. This is a kind of memory of the established network. This has been calculated based on the current input x(t) and the previous-time step's hidden-state h(t-1). Output: The function vector y(t) is the output, at time step t. Weights : Weights: In the RNNs, the input vector-connected to the hidden layer neurons at time t is by a weight matrix of U (Please refer to the above picture). Internally weight matrix W is formed by the hidden layer neurons of time t-1 and t+1, followed by this the hidden-layer with to the output vector y(t) of time t by a V (weight matrix); all the weight matrices U, W, and V are constant for each time step. Advantages Disadvantages It has the special feature that it would remember every each information, so RNN is much useful for time series prediction. The big challenge is during the training period. Perfect for creating complex patterns from the input time series dataset. Expensive computation cost Fast in prediction/forecasting Not affected by missing values, so the cleansing process can be limited I believe this guide would help you all, to understand the time series, flow, and how it works. The media shown in this article is not owned by Analytics Vidhya and are used at the Author's discretion. Related