1. Time series correlation

2019 - 2020

- Import necessary libraries
- import time series data.

Aim:

- to find the rolling correlation of timeseries data

Rolling correlation: It is the relationship between two-time series on a rolling window of a certain size. The rolling window represents a fixed period of time, and as the window rolls forward, the correlation between the two data sets is recalculated for the new window. This allow for the calculation of the correlation between two time series data ~~over~~ sets over time rather than just at one point in time. It can be used to identify trends or patterns in the relationship between two data sets

steps:

- import neccessary libraries
- import time series data
- data preprocessing - remove null values, find percentage change of the columns that we have to compute the correlation.

* set time/date column as index.

- compute correlation between two columns as correlation 1.
- compute rolling correlation over the rolling window 365 days and 720 days as correlation 2 and correlation 3 respectively
- Plot rolling correlation of two columns.

2. Time series ARIMA

Any data recorded with some fixed interval of time is called the time series data. Mainly there are 4 characteristics for time series data. They are Trend (change in dependent variables w.r.t time from start to end), Seasonality (observations that are repeats after fixed time of interval), Irregularities (also called noise, strange dips and jumps in data), cyclic (observations in series repeats in random pattern).

steps :

- import neccessary libraries
- import timeseries data.
- ~~Data preprocessing~~ understanding data and preprocessing : finding time series characteristics, set date column as index, drop duplicates.
- check for the stationality. ie; for accurate analysis and forecasting trend and seasonality should be removed. ie; time series data is said to be stationary when statistical properties like mean, standard deviation should be constant & no seasonality.
- set rolling window = 12 since 12 months in a year.
- for checking stationality, Augmented Dickey fuller test is used.
  p-value <= 0.5 — stationary
  Test statistic — more -ve & should be smaller than critical values (1%., 5%., 10%.)
- If the data is not stationary convert to stationary using

various techniques like differencing, transformation, Moving average, weighted MA etc.

- Since the data is converted to stationary then perform forecasting using ARIMA.

- ARIM : Auto Regressive Integrated Moving Average.
  It is a combination of AR & MA model.

- find AR no compute ACF (auto correlation function) & partial ACF to find the parameters of AR & MA model. They are p & q.

- using the p, q, d parameters find AR & MA
- Then build the combined model. ie; ARIMA
- After model building use it for predictions. After prediction, reverse transformation have to do to the predicted value to get original scale back.
- Then plot the comparison of predicted and actual values
- Next step is to plot the forecasting for next 5 specified years using statmodel plot-predict().
- Print the forecasted value.

3. TF-IDF Recommendation

TF-IDF : Term frequency - inverse document frequency, used to quantify the importance of a word in the document or a set of document.

Steps :

- Import necessary libraries
- load dataset.
- Data preprocessing : removing null value & duplicates
- find the TF using tfidf vectorizer and fit data.
- compute the cosine similarity of the words using tfidf matrix.
- Using cosine similarity generate key, value pair of each word article and its similar words similar to it.
- Using an input word, print the similar words that Similar to its Specific input

4. Time Series LSTM

LSTM : & Long short term memory is a neural network ~~used~~ that handle sequence datas.

Steps :

- Import neceessary libraries
- Load dataset
- Data preprocessing :
- Split dataset into train and test
- Normalise the train and test data using any scaler like MinMaxScaler().
- Define a TimeGenerator and tune its parameters.
- Build a LSTM model and do hyperparameter tunning.
- fit the data to the model and ~~print~~ compute test prediction, and actual prediction (inverse transform of test prediction).
- Plot actual and predicted values to understand the difference between two values.

5. POS tagging and ~~chunking~~ chunking

Pos tagging : labelling each word in a text with its corresponding part of speech

chunking: process of identifying non-overlapping phrases in a sentence.

Steps:

- Import libraries and load to dataset

~~Read the treebank nltk corpus tagged sentences~~

- Read the data with nltk ~~tag~~ universal tagset and print data with its tags
- Define grammer of Noun Phrase
- Define a parser to parse through the tagged data and find the ~~tagged~~ tagset with NP.
- Store all NP ~~np~~ in a list and print the result.