

CS F320 - Foundations of Data Science - Assignment #2

Multivariate Time Series Analytics

Group Members :

1. Akshit Khanna 2017A7PS0023P
2. Vitthal Bhandari 2017A7PS0136P

Tasks :

1) Regression

- a) Identification of an MVTs regression algorithm : We will use one of the most commonly used methods for multivariate time series regression – Vector Auto Regression (VAR). In a VAR model, each variable is a linear function of the past values of itself and the past values of all the other variables.
- b) Identification of an MVTs regression application : Predicting future air quality index values to make informed decisions to tackle the problems. Using our dataset of Air Quality we can predict the future air quality index values based on the past dataset.
- c) Dataset Used : Air Quality Data Set is used for the regression task. The dataset contains 9358 instances of hourly averaged responses with 15 attributes. [Dataset Link](#)

2) Classification

- a) Identification of an MVTs classification algorithm : We will use the k-NN algorithm for classification (with k=1) that uses the dynamic time warping Euclidean distance and PCA similarity factor as the similarity metric. The DWT calculation algorithm will be sped up with the use of the LB Keogh bound and the dynamic time warping locality constraint.
- b) Identification of an MVTs classification application : Predicting whether a subject's eyes are open or closed based only on their brain wave data (EEG). We will discover this problem of predicting based on brain waves and a common methodological trap when evaluating time series forecasting models.
- c) Dataset Used : EEG Eye State Data Set is used for classification for time series. The duration of the measurement is 117 seconds. '1' indicates the eye-closed and '0' the eye-open state. [Dataset Link](#)

3) Clustering

- a) Identification of an MVTs clustering algorithm : We will apply the same idea to clustering that we'll use for classification (using dynamic time warping Euclidean distance and PCA similarity factor as the similarity metric). Different types of clustering such as k-means, k-medoids and hierarchical clustering can be used. In these algorithms, the number of clusters will be set beforehand and similar time series will be clustered together.
- b) Identification of an MVTs clustering application : Clustering of different types of stocks depending on the various values of stocks to interpret the market and for better investment strategy.

- c) Dataset Used : Stock data for S&P 500 companies for a specified length. Cluster them into similar groups.

Work Division :

Implementation of regression algorithm (**Vitthal Bhandari**) ,Application of regression algorithm (**Akshit Khanna**), Implementation of classification algorithm (**Vitthal Bhandari**), Application of classification algorithm (**Vitthal Bhandari**), Implementation of clustering algorithm (**Akshit Khanna**), Application of clustering algorithm (**Akshit Khanna**).

All the implementations will be done in **R**.