Time Series Project

Jacob Clinton George Smith-Kolff

2024-09-22

Contents

Chapter 1

Introduction

(Probably work on this section last)

1.1 What is missing data

1.2 Issues caused by missing data

1.3 Structure of the report

Chapter 2

Theory

2.1 Time series

We begin by providing definitions for both uni variate and multivariate time series:

A univariate time series $X = \{x_1, x_2, ..., x_t\} \in \mathbb{R}^t$ is a sequence t observations on a single variable.
This can be extended to multivariate time series:
$X = \{x_1, x_2, ..., x_t\} \in \mathbb{R}^{t \times d}$ where each x_i is a d dimensional vector.

2.2 Missing data

remove the subsubsection numbers later

Data missingness is a common occurrence that is common in working with real world data. Missing data can arise from device failures such as measureing equipment failing, or from data censoring (such as from governments) [1]. Missing data typically falls into one of three categories:

2.2.1 Missing Completely at Random (MCAR)

Data is said to be missing completely at random if the distribution that describes how missingness occurs is independent of both the observed and unobserved values in the time series.

2.2.2 Missing at Random (MAR)

Missing at random is when missingness is related to the observed data, but is independent of the unobserved data. This means that there is some external factor that is causing the missingness. An example given in [2] is that data from a sensor is more likely to be missing on weekends since on some weekends the sensor is shutdown.

2.2.3 Not Missing at Random (NMAR)

Not missing at random means that the missingness is related to the value of the observation itself. An example of this is a sensor that will return a missing value if the recorded value is above 100°.

Come back and include probability notation

2.3 Missing data imputation

2.3.1 Multivariate data imputation

2.3.1.1 Multivariate data imputation techniques

2.3.1.1.1 K-Nearest Neighbours

2.3.1.1.2 Multivariate Imputation by Chained Equations (MICE)

2.3.1.1.3 General Adversial Networks (GAM)

2.3.2 Univariate data imputation

2.3.2.1 The struggle with univariate data imputation

2.3.2.2 Last Observed Carried Forward (LOCF) Next Observation Carried Backward (NOCB)

2.3.2.3 Mean/Median imputation

2.3.2.4 Kalman filtering

2.3.2.5 Linear Interpolation

Chapter 3

Illustration of the method application

Using the methods of missing data in practice.

3.1 The data

3.2 The methods

Chapter 4

Conclusion

[1] R. J. Little and D. B. Rubin, *Statistical analysis with missing data*, vol. 793. John Wiley & Sons, 2019.

[2] S. Moritz, A. Sardá, T. Bartz-Beielstein, M. Zaefferer, and J. Stork, "Comparison of different methods for univariate time series imputation in r," *arXiv preprint arXiv:1510.03924*, 2015.

Chapter 5

Appendix