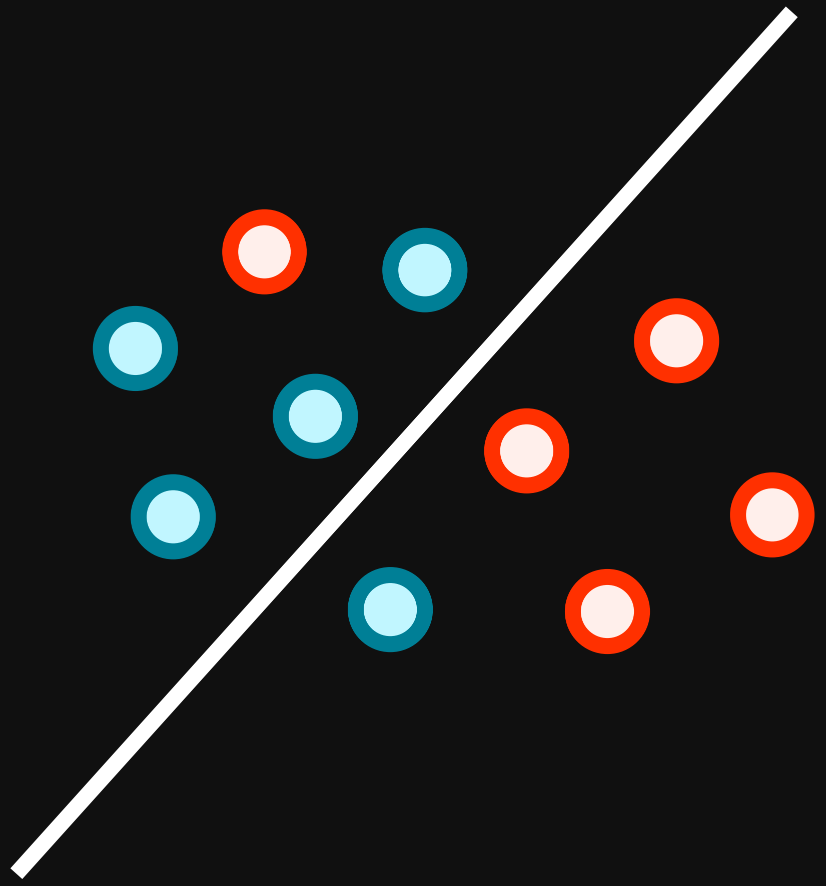
Basics of Machine Learning

Dmitry Ryabokon, github.com/dryabokon



Lesson 18
Time Series

Benchmarking

Summary

- Key definitions
  - Stationarity, Trends, Seasonality
  - Univariate vs Multivariate
  - Exogeneous variables
- Approaches
- Gated Recurrent Units (GRU)

https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/

https://github.com/jbrownlee/Datasets

https://www.kaggle.com/prashant111/complete-guide-on-time-series-analysis-in-python

https://www.kaggle.com/kashnitsky/topic-9-part-1-time-series-analysis-in-python

Key definitions

Key definitions

Univariate vs Multivariate time series

Univariate time series

- Only one variable is varying over time.
- For example, data collected from a sensor measuring the temperature of a room every second. Therefore, each second, we have a one-dimensional value, which is the temperature.

Multivariate time series

- Multiple variables are varying over time.
- For example, a tri-axial accelerometer. There are three accelerations, one for each axis (x,y,z) and they vary simultaneously over time.

Key definitions

Endogenous vs Exogenous variables

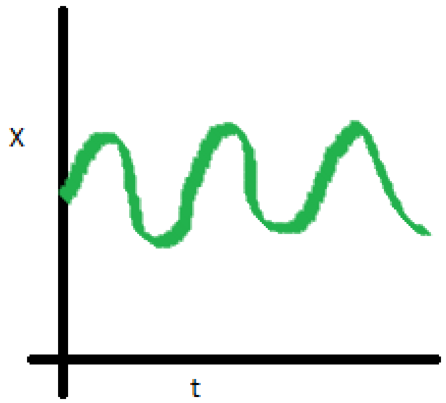
An input variable is **endogenous** if it is affected by other variables in the system and the output variable depends on it.

An input variable is an **exogenous** if it is independent of other variables in the system and the output variable depends upon it.

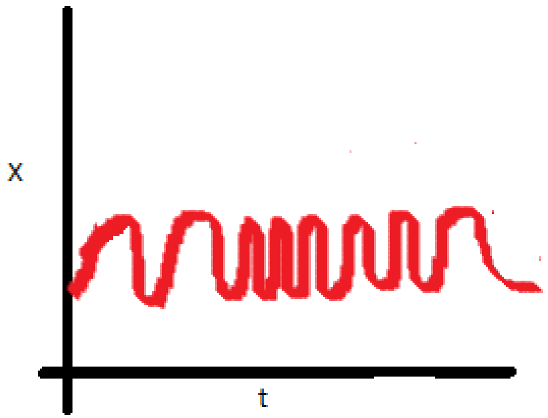
Typically, a time series forecasting problem has endogenous variables and may or may not have exogenous variables.

Key definitions

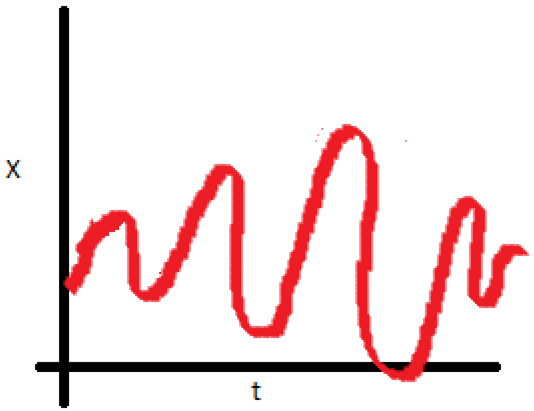
Stationarity



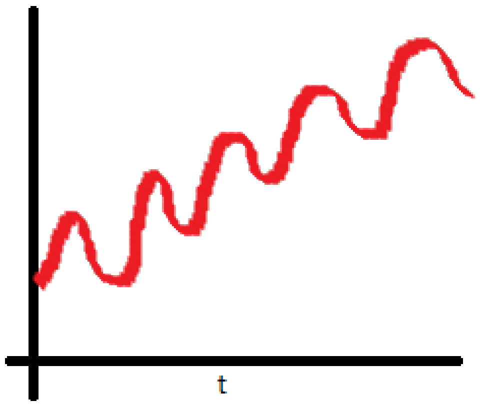
Stationary series



Non-Stationary series



Non-Stationary series



Non-Stationary series

Key definitions

Stationarity

Reasons to convert a non-stationary series into a stationary one

- Forecasting a stationary series is relatively easy and the forecasts are more reliable.
- An important reason is, autoregressive forecasting models are essentially linear regression models that utilize the lag(s) of the series itself as predictors.
- We know that linear regression works best if the predictors (X variables) are not correlated against each other. So, stationarizing the series solves this problem since it removes any persistent autocorrelation, thereby making the predictors(lags of the series) in the forecasting models nearly independent.

Key definitions

Seasonality

Additive time series:

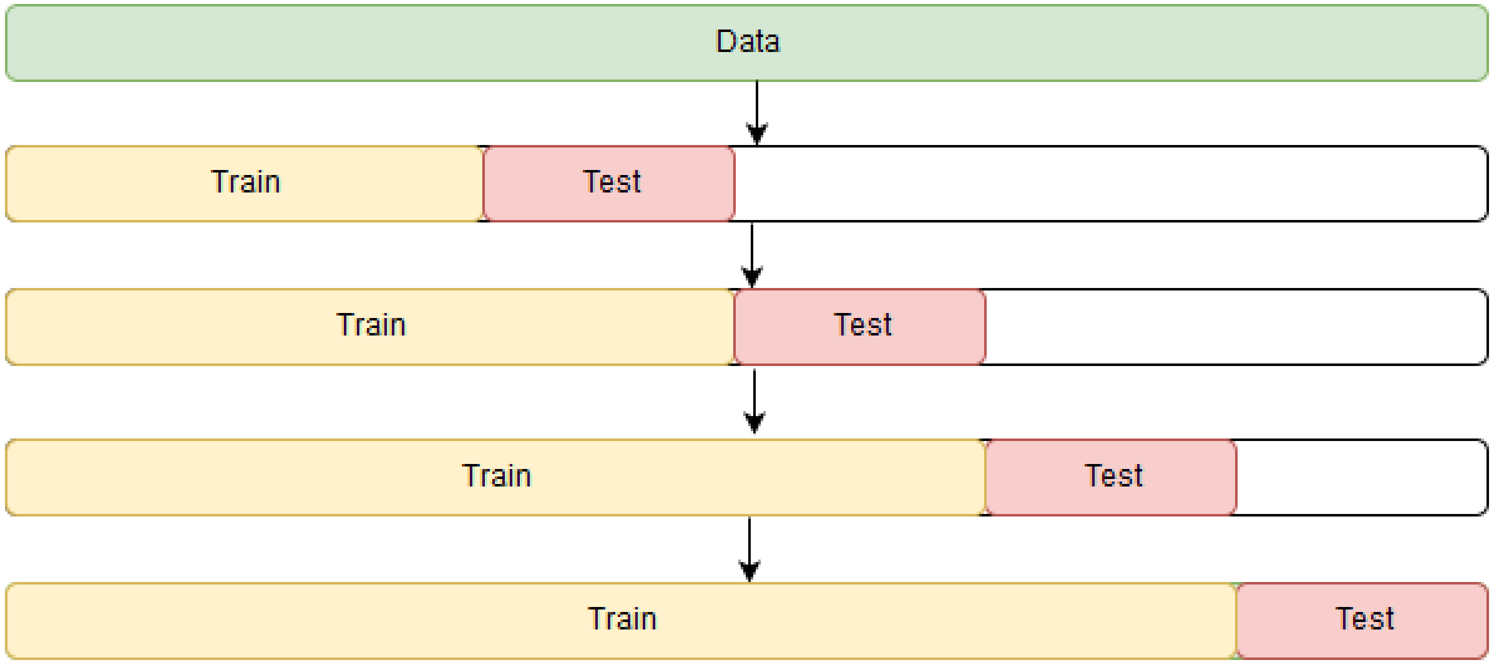
- Value = Base Level + Trend + Seasonality + Error

Multiplicative Time Series:

- Value = Base Level x Trend x Seasonality x Error

Key definitions

Cross Validation for Time Series



Forecasting approaches

Forecasting Approaches

Classic approaches

	Multivariate	Trend	Seasonality	Exogenous vars
Autoregression (AR)	-	-	-	-
Moving Average (MA)	-	-	-	-
Autoregressive Moving Average (ARMA)	-	-	-	-
Autoregressive Integrated Moving Average (ARIMA)	-	+	-	-
Seasonal Autoregressive Integrated Moving-Average (SARIMA)	-	+	+	-
SARIMA with Exogenous Regressors (SARIMAX)	-	+	+	+
Vector Autoregression (VAR)	+	-	-	-
Vector Autoregression Moving-Average (VARMA)	+	-	-	-
Vector Autoregression Moving-Average with Exogenous Regressors (VARMAX)	+	-	-	+
Simple Exponential Smoothing (SES)	-	-	-	-
Holt Winter’s Exponential Smoothing (HWES)	-	+	+	-

https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/

Forecasting Approaches

Links

https://medium.com/analytics-vidhya/text-lstm-f1aaceeb5727

https://people.duke.edu/~rnau/411arim.htm

https://people.duke.edu/~rnau/411arim3.htm

https://www.digitalocean.com/community/tutorials/a-guide-to-time-series-forecasting-with-arima-in-python-3

https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21

https://colah.github.io/posts/2015-08-Understanding-LSTMs/

https://github.com/PsiPhiTheta/LSTM-Attention/blob/master/report/main.pdf