Time Series Forecasting

2023/01
First Week of 2023 Study
Jiwon Kim

Types of Time Series Forecasting

1. Univariate Time Series
- Single Variable observation sequence over time
- E.g. Historical temperature for a given region
1. Multivariate Time Series
- Multiple variable observation sequences over time
    - The variables can be dependent to each other (time-effects and other interactions)
        - Exogenous and Endogenous
            - Exogenous: input variable not influenced by others
            - Endogenous: input variable influenced by others

Time Series Forecasting Methods

1. Classical/ Statistical Models
    a. Moving Average, Exponential Smoothing, ARIMA, SARIMA, TBATS
2. Machine Learning
    a. Linear Regression, XGBoost, Random Forest, or any ML model with reduction methods
3. Deep Learning
    a. RNN, LSTM

Statistical Models

ARIMA (Auto-regressive Integrated Moving Average)

Characteristics

1. Autoregression (AR) : Model that shows changing variable that regresses on its own previous values
2. Integrated (I): differencing of raw observations to allow for the time series to become stationary (|Y| = 0)
    a. Differencing between current and the past has been taken at least once
3. Moving Average: incorporating the dependency between observation and residual error from moving average model applied to logged observations
    a. Regression error actually be expressed in linear combination of error series

Statistical Models

SARIMA (ARIMA w/ seasonal component)

Characteristics

1. ARIMA + hyperparameter that deals with seasonal component, namely “repeating cycle”

Statistical Models

Exponential Smoothing: Forecasting for univariate time series data that supports data with systemic trend or seasonal component. Popular alternative to ARIMA

Methods Family

1. vSimple (single): Weighted moving average with exponentially decreasing weights
2. Double Exponential Smoothing: vSimple + beta smoothing factor that controls the decay of trend change
3. Triple Exponential Smoothing: Double Exponential Smoothing + gamma that controls the influence of seasonal component

Statistical Models

Exponential Smoothing: Forecasting for univariate time series data that supports data with systemic trend or seasonal component. Popular alternative to ARIMA

Methods Family : vSimple

1. Controlled by a single parameter (smoothing factor) called alpha

Statistical Models

Exponential Smoothing: Forecasting for univariate time series data that supports data with systemic trend or seasonal component. Popular alternative to ARIMA

Methods Family: Double Exponential Smoothing

1. Holt's Smoothing: Double Exponential Smoothing w/ linear trend
2. Multiplicative Trend: Double Exponential Smoothing w/ exponential trend

Statistical Models

Exponential Smoothing: Forecasting for univariate time series data that supports data with systemic trend or seasonal component. Popular alternative to ARIMA

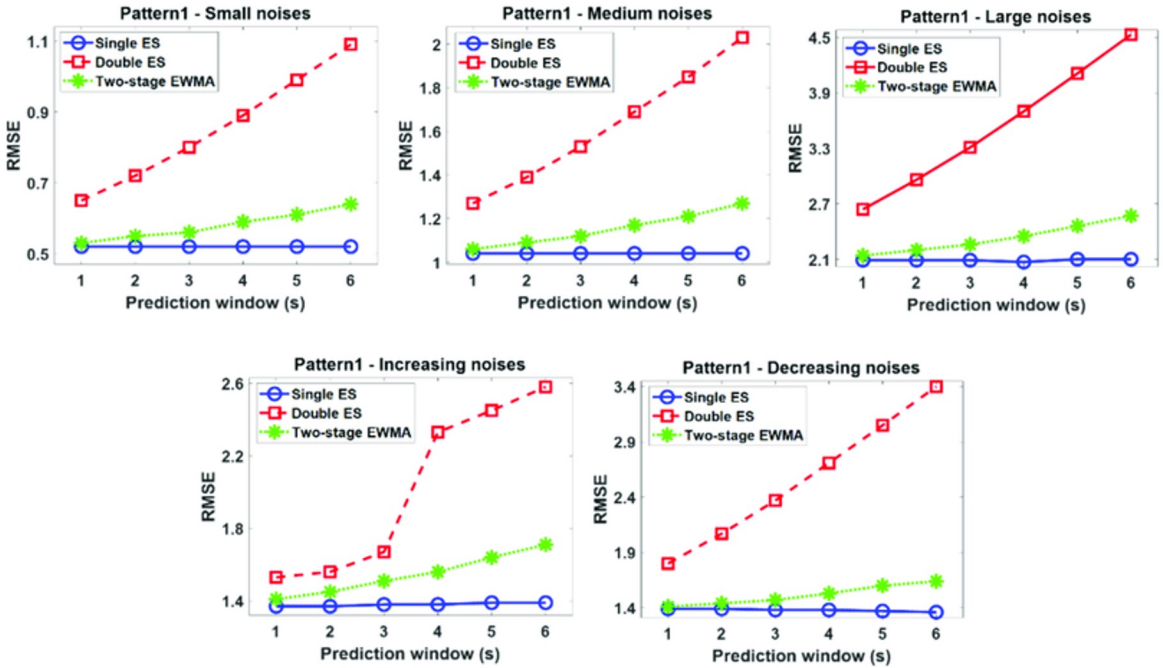
Methods Family: Triple Exponential Smoothing

1. Additive Seasonality: Triple Exponential Smoothing with a linear seasonality.
2. Multiplicative Seasonality: Triple Exponential Smoothing with an exponential seasonality

Hyperparameters

- **Alpha** (α): Smoothing factor for the level.
- **Beta** (β): Smoothing factor for the trend.
- **Trend Type**: Additive or multiplicative.
- **Dampen Type**: Additive or multiplicative.
- **Phi** (ϕ): Damping coefficient.
- **Gamma** (γ): Smoothing factor for the seasonality.
- **Seasonality Type**: Additive or multiplicative.
- **Period**: Time steps in seasonal period.

Statistical Models: Comparison results of Exponential Smoothing



Statistical Models: TBATS

Time Series with multiple seasonality (e.g. daily, weekly pattern)

1. Box-Cox transformation is applied to the original
2. Modeled as a linear combination of exponentially smoothed trend, seasonal and ARMA components

Cite:
https://www.sktime.org/en/v0.15.1/api_reference/auto_generated/sktime.forecasting.tbats.TBATS.html#re7d54b10b246-1

Machine Learning

Automated Machine Learning Tools (autoML) that let you compare multiple models at once:

(1) Autogluon (2) Py-caret (3) auto-sklearn (4) MLBox (5) TPOT (6) H2O (7) auto-Keras

- If using cross-validation, must use time-series appropriate techniques
  - Train data should always be formed with observations prior to the observations that form the test set
  - Pycaret uses TimeSeriesSplit from the scikit-learn library

https://www.automl.org/automl/
https://www.linkedin.com/pulse/top-10-automated-machine-learningauto-ml-tools-used-2020-2021-sahu/

Notable ML Forecasting Framework

1. Meta’s Open-Source - Prophet
2. Sklearn
3. Pmdarima
4. Meta’s Open-source - Kats
5. Uber’s Open-source- Uber

Notable ML Forecasting Framework - Prophet

Univariate time analysis:

Regression Trend component: a piecewise trend because it is made from data that has been broken into pieces using knots.

Seasonal component: Fourier variables

Point (Pulse) component: Binary variable, like indicating Holiday, or other possibly contributing factors

Prophet description:
https://www.youtube.com/watch?v=2XFro0nIHQM

Onto Deep Learning

CNN, LSTM

Reference

https://www.datacamp.com/tutorial/tutorial-time-series-forecasting

Predict the Future with MLPs, CNNs and LSTMs in Python, Jason Brownlee, Machine Learning Mastery

w/ list of links that are in the speaker notes