



CIS 635 - Knowledge Discovery & Data Mining

Time series data modeling



Outline

- How are the time series problems different than non time series problems
- Stationary vs non-stationary signals
- Signal decomposition
- AR(I)MA: Autoregressive Integrated Moving Average
 - From non no-stationary to stationary
 - Auto Regression
 - Moving Average



How are the time-series problems different?

- The models (Regression and Classification), we have learned so far are of the form:

$$f(y|X)$$

- For certain data, especially the time series, we can take advantage of the form:

$$f(y_t|X, y_{t-1}, y_{t-2}, \dots, y_0); \text{ essentially, here the input features are } X \text{ plus the lagged instances of the target } y.$$

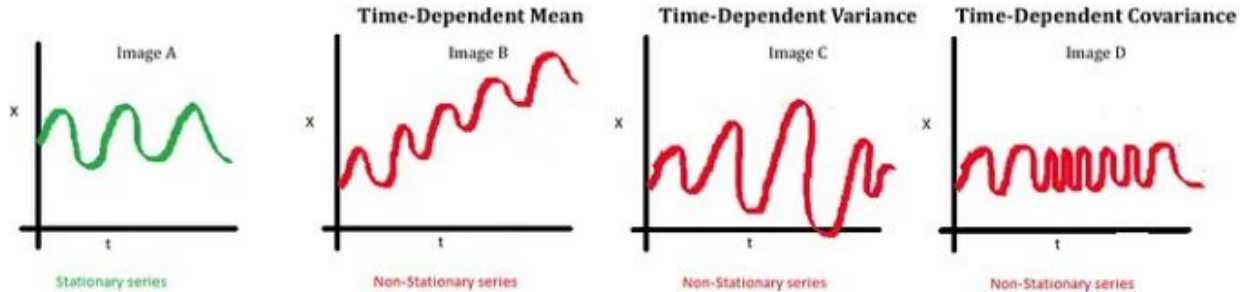
- Purely time series models are of form :

$$f(y_t|y_{t-1}, y_{t-2}, \dots, y_0), \text{ where there is no explicit, } X.$$

Stationarity vs non-stationary signals

- A time-series is said to be stationary if it does not display any trends or seasonality.
- One more way of defining stationarity is that it is when data does not have any time-dependent mean, variance or covariance.

The Principles of Stationarity



[ref](#)



Non stationary to stationary

- If signal is non-stationary, we can convert them into stationary signal by differencing

$$T_t = S_t - S_{t-1},$$



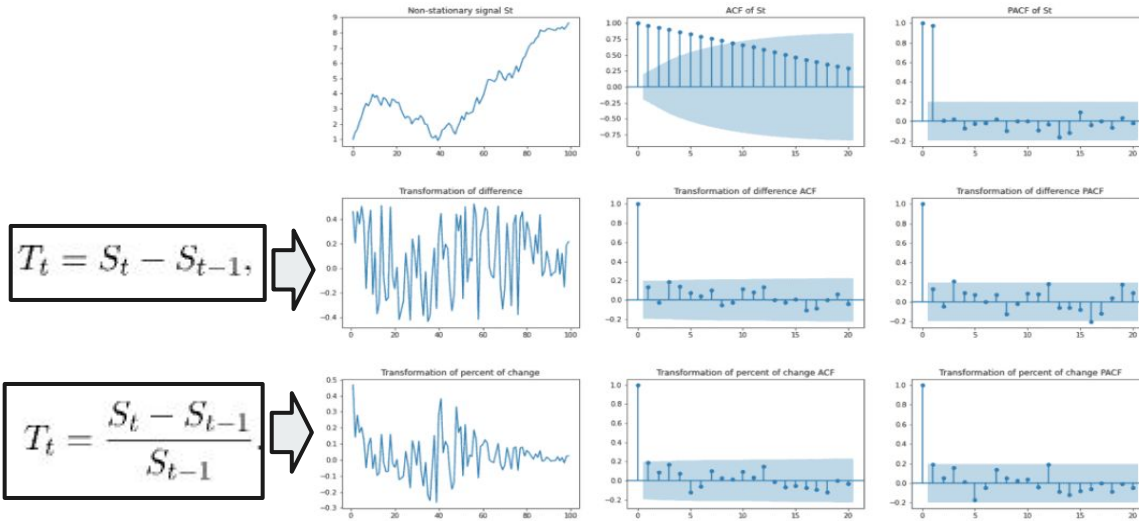
Non stationary to stationary

- If signal is non-stationary, we can convert them into stationary signal by differencing
- or calculating percent of change

$$T_t = S_t - S_{t-1},$$

$$T_t = \frac{S_t - S_{t-1}}{S_{t-1}}.$$

Non stationary to stationary





Statistical time series models

- Moving Average (MA)
- Autoregressive Models (AR)
- Autoregressive Moving Average (ARMA)
- Autoregressive Moving Integrated Average (ARIMA)

$$f(y_t | y_{t-1}, y_{t-2}, \dots, y_0)$$



AR(I)MA

- AutoRegressive Integrated Moving Average (ARIMA) is a statistical model for forecasting time series data.

A combined model with **AR**, **MA**, but first transforming the signal to stationary.

- **AR (Autoregression):** This emphasizes the dependent relationship between an observation and its preceding or 'lagged' observations.
- **I (Integrated):** To achieve a stationary time series, one that doesn't exhibit trend or seasonality, differencing is applied. It typically involves subtracting an observation from its preceding observation.
- **MA (Moving Average):** This component zeroes in on the relationship between an observation and the residual error from a moving average model based on lagged observations.



AR(I)MA

The parameters of the ARIMA(p,d,q) model are defined as follows:

- **p**: The lag order, representing the number of lag observations incorporated in the model.
- **d**: Degree of differencing, denoting the number of times raw observations undergo differencing.
- **q**: Order of moving average, indicating the size of the moving average window.

A combined model with **AR**, **MA**, but first transforming the signal to stationary.

[ref](#)



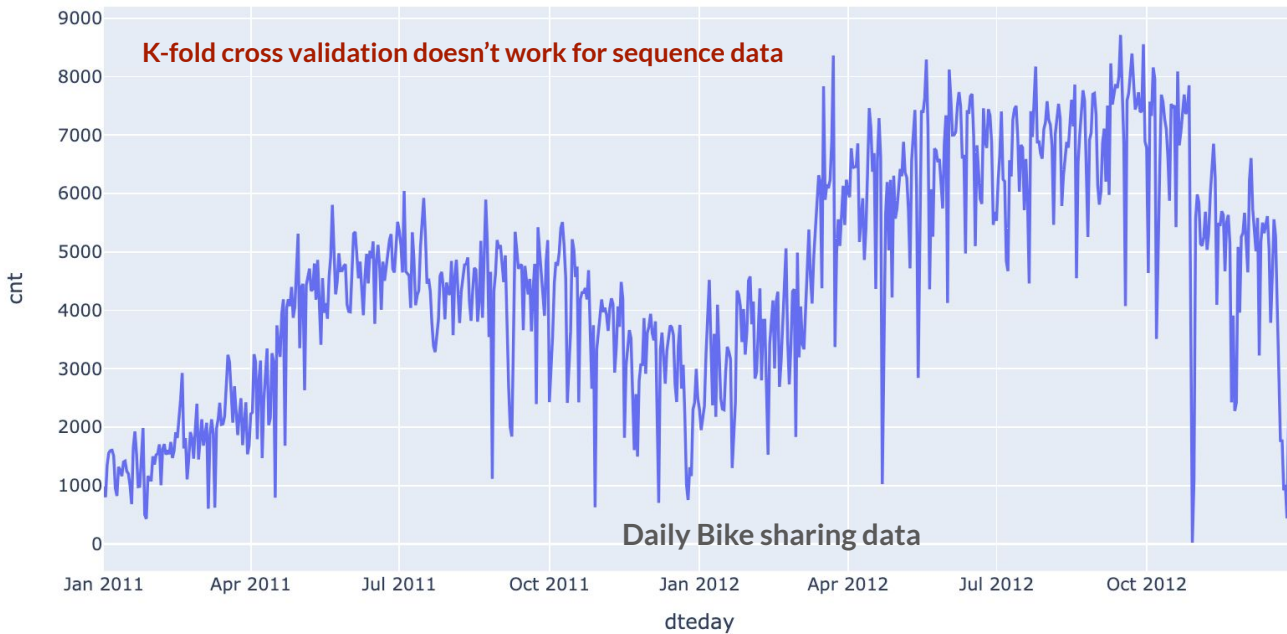
Notebook presentation

Notebook

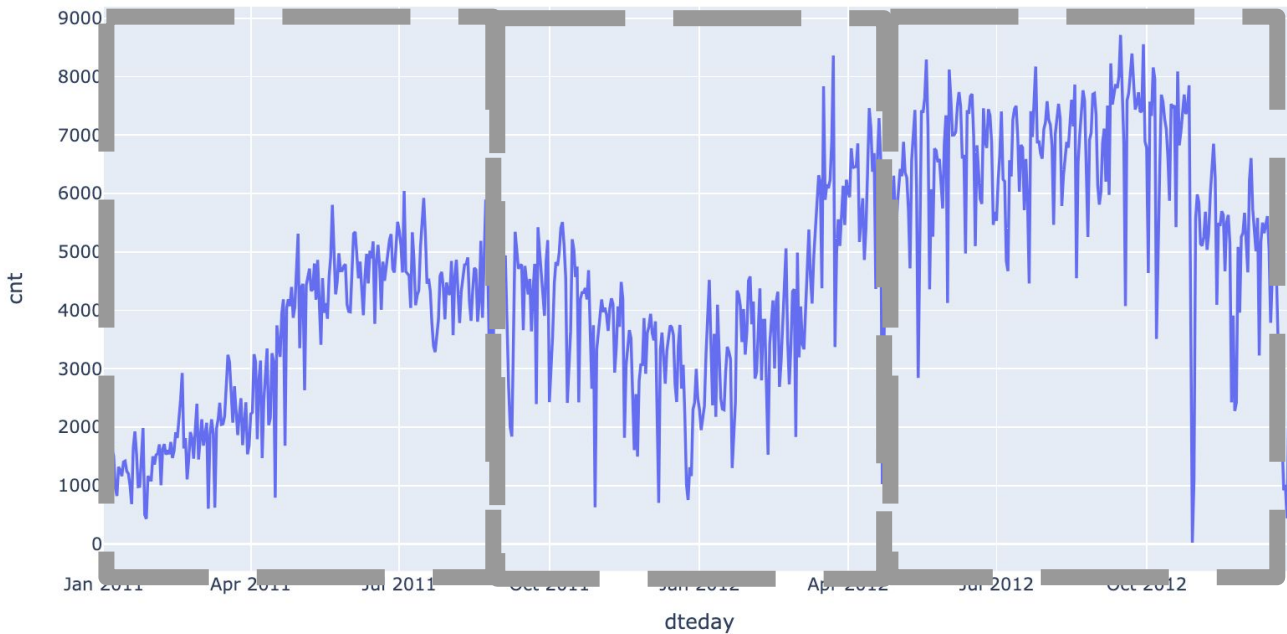


Sequential cross validation

Sequential cross validation



Sequential cross validation



Sequential cross validation



Valid configuration:

Training on **first two folds**, and test on the **last fold**

Sequential cross validation

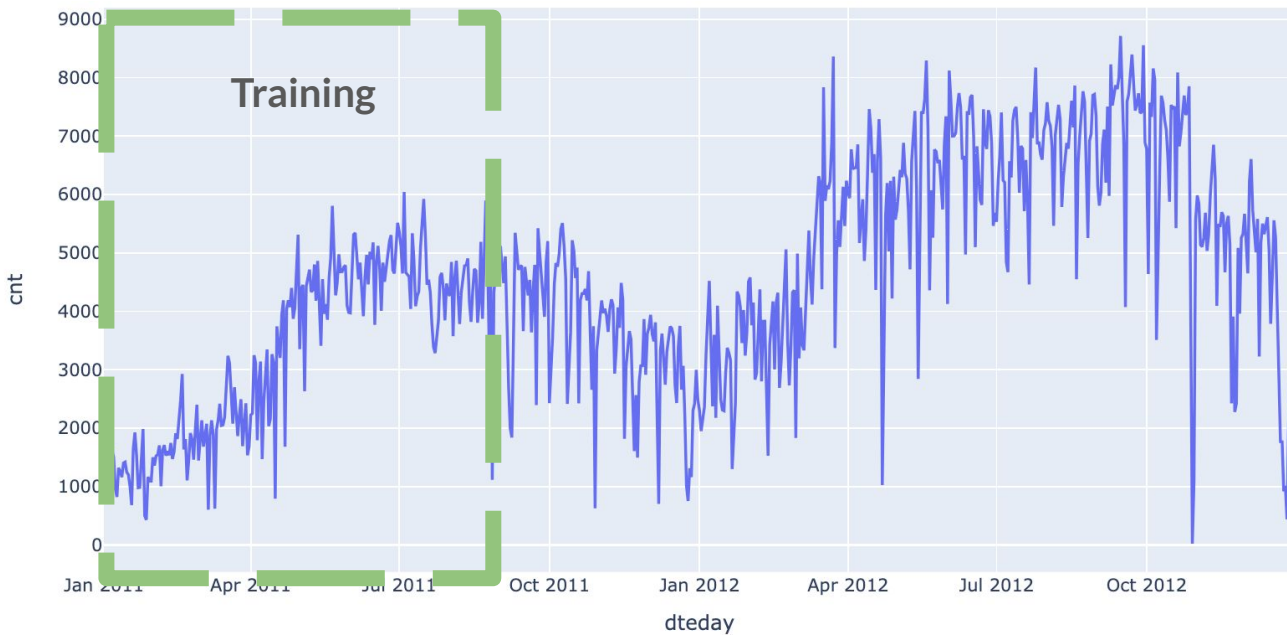


*Invalid
configuration:
Training on **first
fold**, and test on the
last two.*



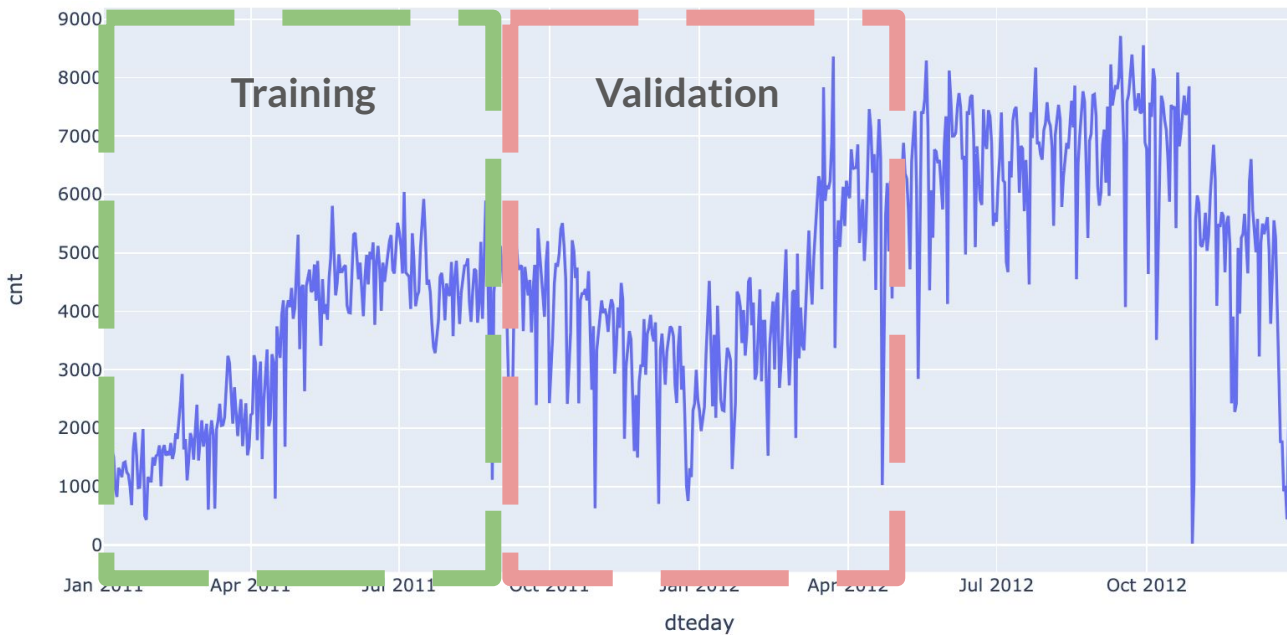
Sequential cross validation

Sequential cross validation



Always follow!

Sequential cross validation



Always follow!

Sequential cross validation



Always follow!

Sequential cross validation



Always follow!



QA