Introduction to Train and Validation Set

Why use a Validation set?

- ✓ Historical Data
- ✓ Train the model

Analytics Vidhya

Why use a Validation set?

✓ Historical Data

✓ Train the model

x Evaluate the model

Analytics Vidhya

Why use a Validation set?

Analytics Vidhya

Why use a Validation set?

- Historical Data
- Train the model

- Historical Data
- Not used for training
- Evaluate model

Analytics Vidhya

Why use a Validation set?

Analytics Vidhya

Validation Technique for Time Series

- Time series data is sequential
- The ordering of data points is important

Validation Technique for Time Series

- Time series data is sequential
- The ordering of data points is important

Note: Do not shuffle the data

Validation Technique for Time Series

- Time series data is sequential
- The ordering of data points is important

Note: Do not shuffle the data

- Popular Validation techniques:
  - Hold out validation
  - Time Series Cross Validation
  - Walk forward cross validation

Analytics Vidhya

Hold-Out Validation Method

Date	Variable
Day1	y_1
Day2	y_2
.	.
.	.
Dayn	y_n
Dayn+1	y_{n+1}
Dayn+2	y_{n+2}
. .	. .

Analytics Vidhya

Hold-Out Validation Method

Date	Variable
Day1	y_1
Day2	y_2
.	.
.	.
Dayn	y_n
Dayn+1	y_{n+1}
Dayn+2	y_{n+2}
. .	. .

Training Data

Validation Data

Analytics Vidhya

Hold-Out Validation Method

- Drawbacks :
  - A single validation set
  - Model evaluated only once

Hold-Out Validation Method

- Drawbacks :
  - A single validation set
  - Model evaluated only once
- Other validation techniques
  - Time Series Cross Validation
  - Walk forward cross validation

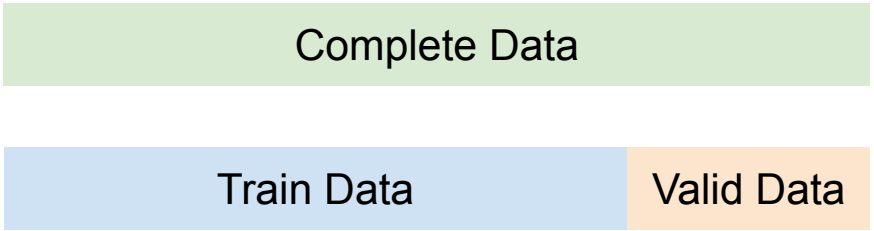
Validation Techniques for Time Series

- Hold-out Validation
- Time Series Cross Validation
- Walk Forward Cross Validation

Analytics Vidhya

Cross Validation for Time Series

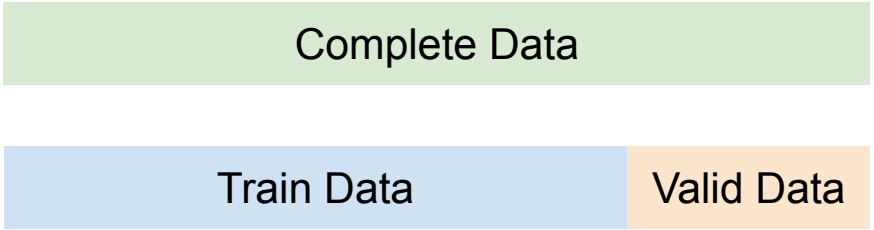
Hold-Out Validation Method



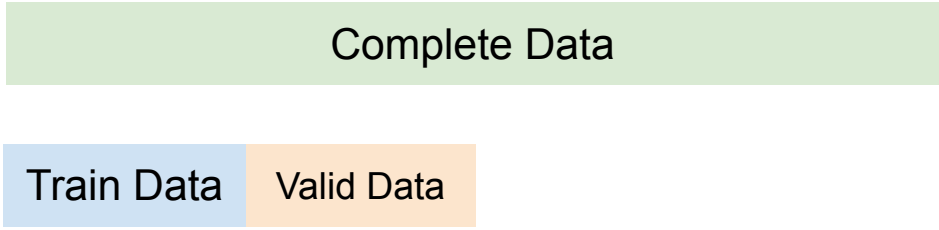
Analytics Vidhya

Cross Validation for Time Series

Hold-Out Validation Method

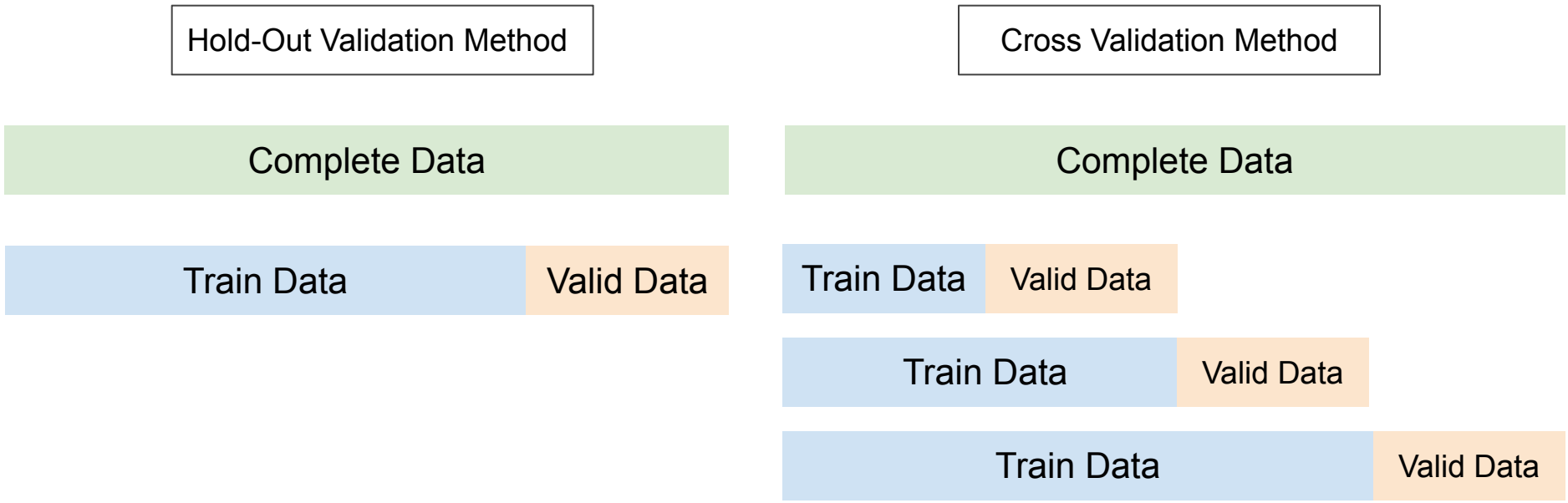


Cross Validation Method



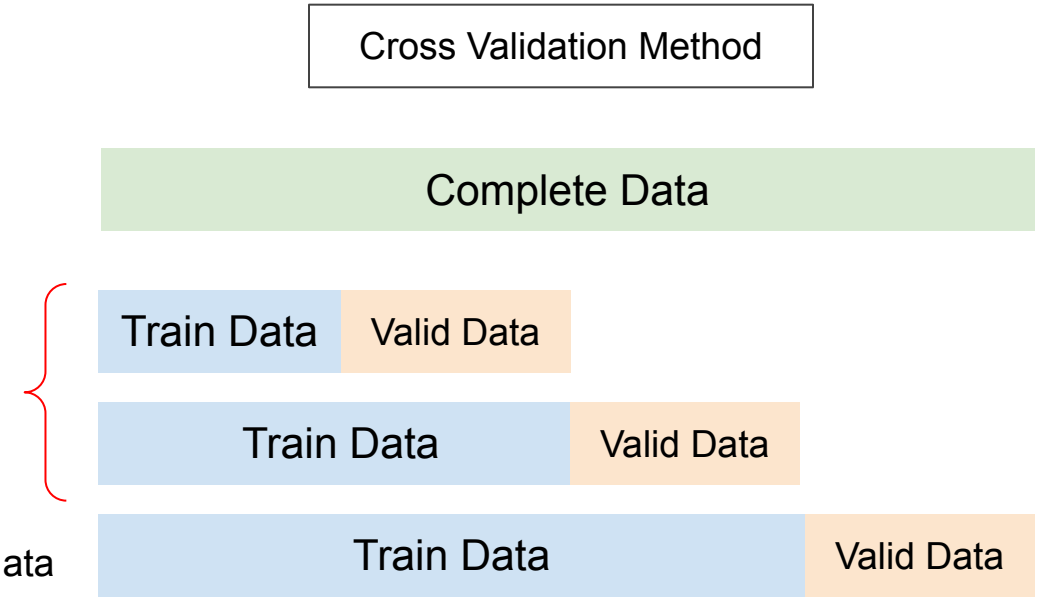
Analytics Vidhya

Cross Validation for Time Series



Analytics Vidhya

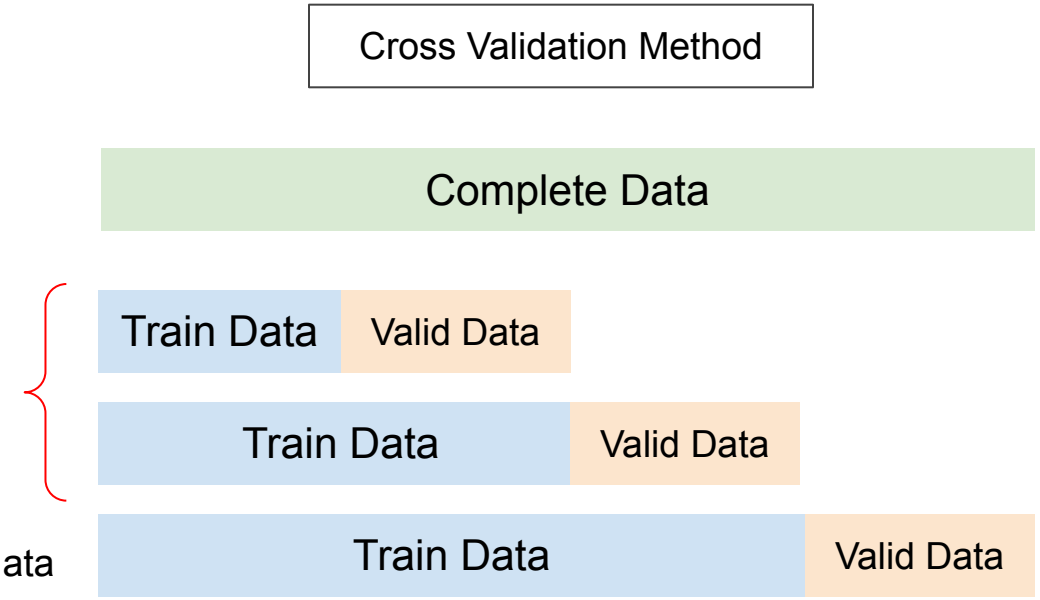
Cross Validation for Time Series



- How well the model performs in past
- How well the model will perform for recent data

Analytics Vidhya

Cross Validation for Time Series



- How well the model performs in past
- How well the model will perform for recent data

Note: Training Data should not include information from the future.

Analytics Vidhya

Validation Techniques for Time Series

- Advantages
  - Model Evaluated on multiple validation sets
  - Average scores across validation sets
- Drawbacks
  - Varying length of train data

Validation Techniques for Time Series

- Hold-out Validation
- Cross Validation
- Walk Forward Cross Validation

Analytics Vidhya

Walk Forward Cross Validation for Time Series

Cross Validation Method

Complete Data

Train Data | Valid Data

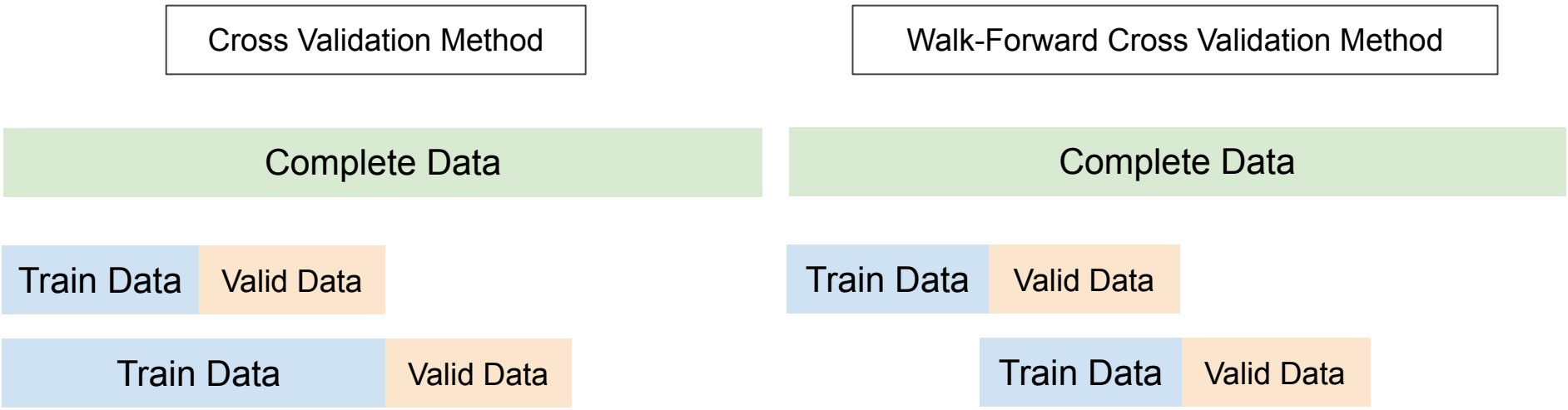
Walk-Forward Cross Validation Method

Complete Data

Train Data | Valid Data

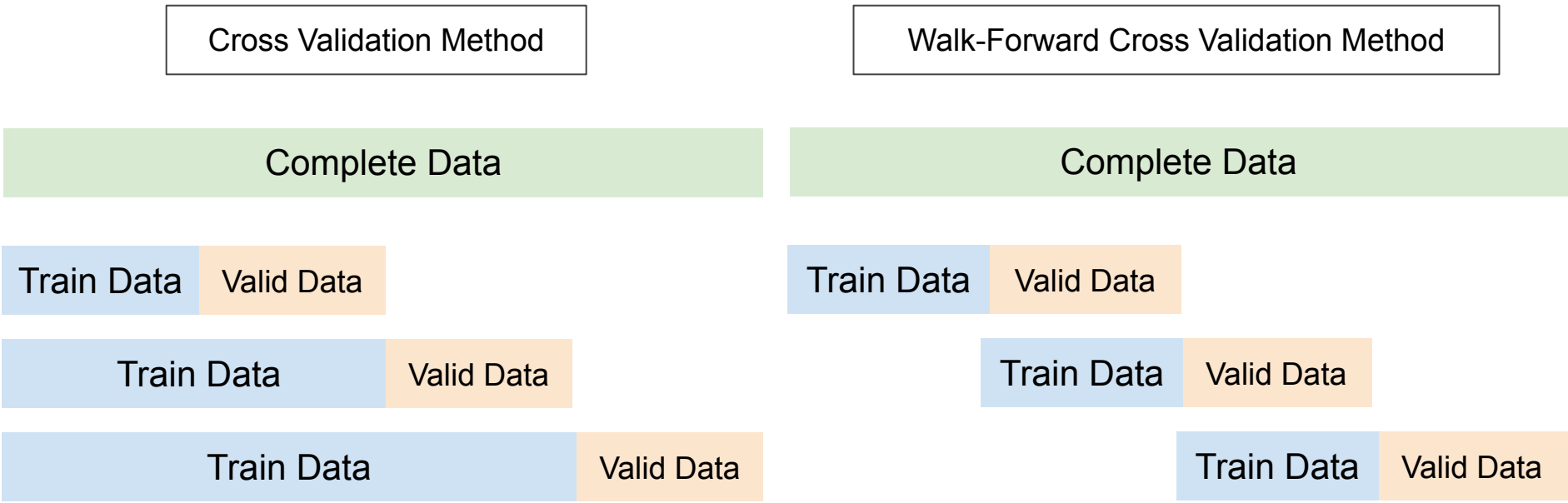
Analytics Vidhya

Walk Forward Cross Validation for Time Series



Analytics Vidhya

Walk Forward Cross Validation for Time Series



Analytics Vidhya

Notebook

When to not use Cross Validation?

- Other Business factors have drastically changed in the past

Overview of the Module

- Define the problem statement
- Evaluation Metrics for time series
- Validation Techniques
- Feature Extraction
- Build ML model

Analytics Vidhya

Overview of the Module

- Define the problem statement
- Evaluation Metrics for time series
- Validation Techniques
- Feature Extraction
- Build ML model

Analytics Vidhya