Linear Regression for Time Series

Analytics Vidhya

Steps Involved in Forecasting Tasks

1. Defining the Problem Statement ✔
2. Data Exploration ✔
3. Data Preprocessing ✔
4. Selecting Evaluation Metric ✔
5. Building & evaluating Models

Analytics Vidhya

Data Preprocessing in Time Series

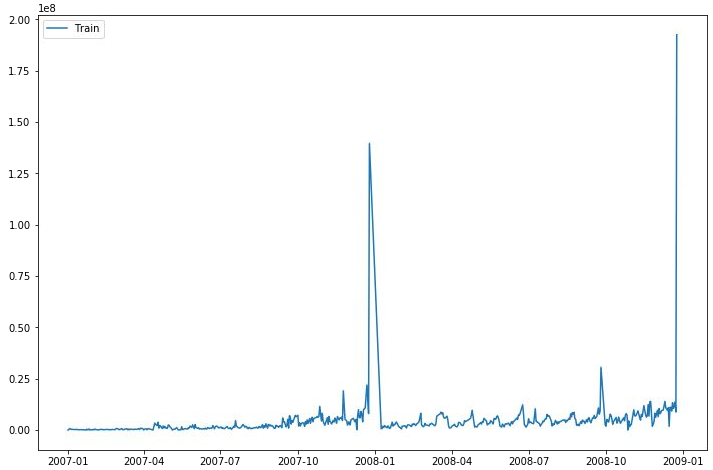
- Does the data have missing values?

	Date	count
21	2012-09-15	3.0
22	2012-09-16	2.0
23	2012-09-17	NaN
24	2012-09-18	7.0
25	2012-09-19	6.0

Analytics Vidhya

Data Preprocessing in Time Series

- Does the data have missing values?
- Are there any extreme values?



Analytics Vidhya

Data Preprocessing in Time Series

- Does the data have missing values?
- Are there any extreme values?
- Are any dates skipped from the series?

Date	Unit_sold
2008-11-20	12416871
2008-11-22	13726524
2008-11-23	13947472

Analytics Vidhya

Data Preprocessing in Time Series

Analytics Vidhya

Data Preprocessing in Time Series

- Error in the data collection process
- Missing/Extreme values on purpose
- Missing Values at random
- Data not captured on Holidays

Analytics Vidhya

Data Preprocessing in Time Series

- Previously we used mean/median/mode

Analytics Vidhya

Data Preprocessing in Time Series

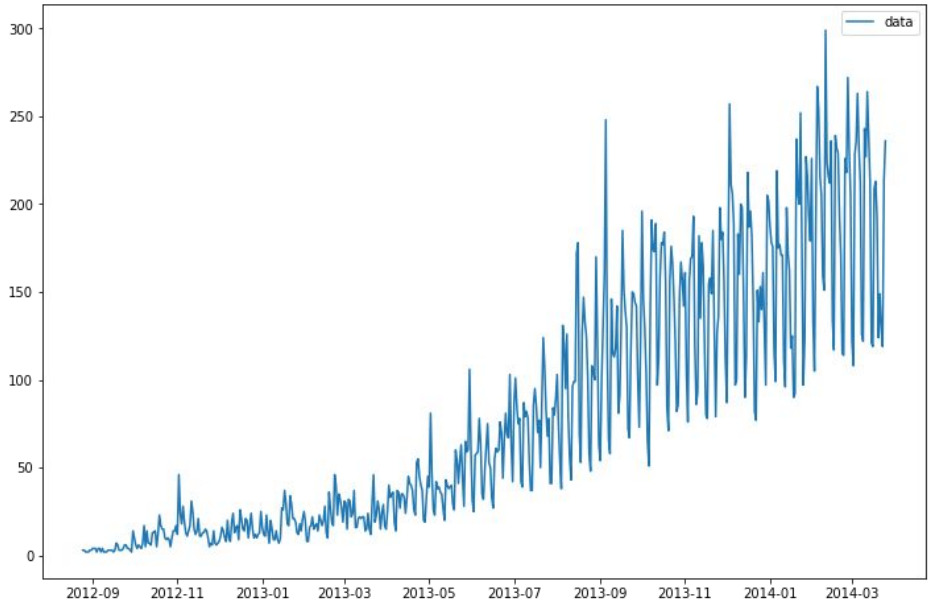
- Previously we used mean/median/mode
- In time series, subsequent observations depend on the previous values

Analytics Vidhya

Data Preprocessing in Time Series

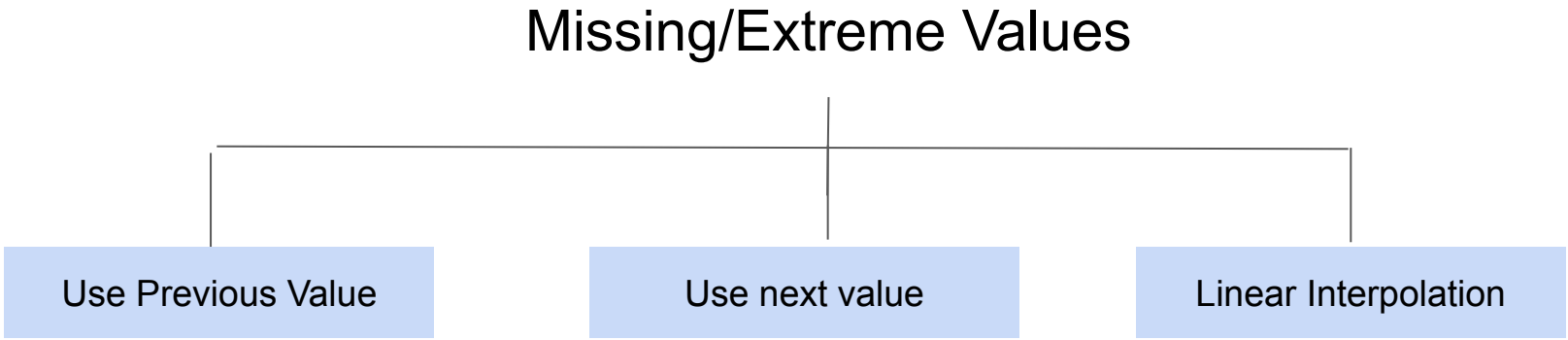
- Previously we used mean/median/mode
- In time series, subsequent observations depend on the previous values

Date	#count
2013-09-24	150
2013-09-25	149
2013-09-26	144
2013-09-27	142
2013-09-28	101
2013-09-29	73
2013-09-30	123
2013-10-01	196
2013-10-02	146
2013-10-03	129
2013-10-04	104
2013-10-05	65
2013-10-06	51
2013-10-07	142
2013-10-08	191
2013-10-09	174



Data Preprocessing in Time Series

- Previously we used mean/median/mode
- In time series, subsequent observations depend on the previous values



Analytics Vidhya

Data Preprocessing in Time Series

Missing/Extreme Values

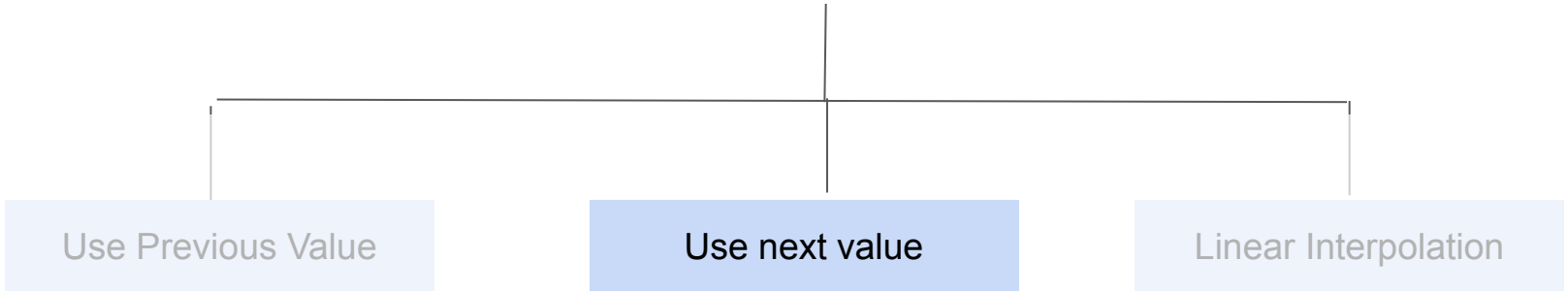
- Use Previous Value
- Use next value
- Linear Interpolation

	Date	count
21	2012-09-15	3.0
22	2012-09-16	2.0
23	2012-09-17	NaN
24	2012-09-18	7.0
25	2012-09-19	6.0

Analytics Vidhya

Data Preprocessing in Time Series

Missing/Extreme Values

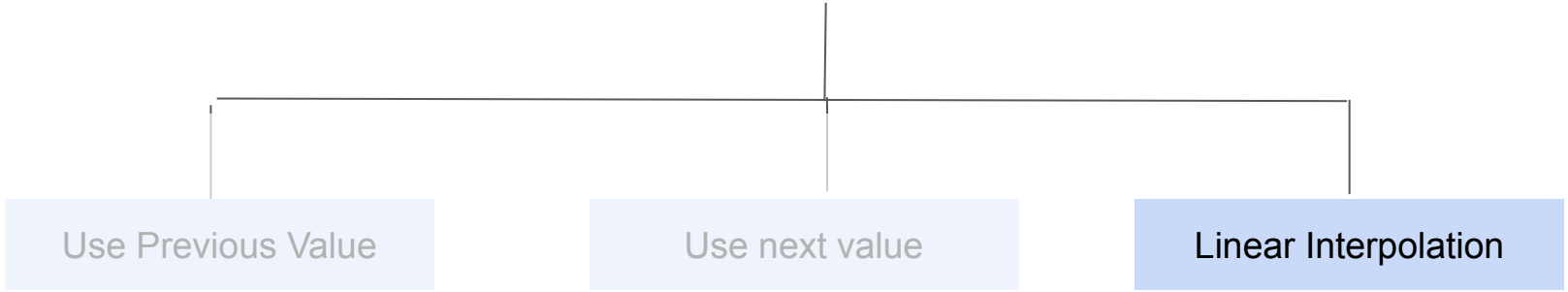


	Date	count
21	2012-09-15	3.0
22	2012-09-16	2.0
23	2012-09-17	NaN
24	2012-09-18	7.0
25	2012-09-19	6.0

Analytics Vidhya

Data Preprocessing in Time Series

Missing/Extreme Values



	Date	count
21	2012-09-15	3.0
22	2012-09-16	2.0
23	2012-09-17	NaN
24	2012-09-18	7.0
25	2012-09-19	6.0

Data Preprocessing in Time Series

S. No.	Date	Count
31	2012-08-04	119
32	2012-08-05	207
33	2012-08-06	201
34	2012-08-07	NaN
35	2012-08-08	NaN
36	2012-08-09	NaN
37	2012-08-10	229

Analytics Vidhya

Data Preprocessing in Time Series

S. No.	Date	Count
31	2012-08-04	119
32	2012-08-05	207
33	2012-08-06	201
34	2012-08-07	NaN
35	2012-08-08	NaN
36	2012-08-09	NaN
37	2012-08-10	229

Analytics Vidhya

Data Preprocessing in Time Series

S. No.	Date	Count
31	2012-08-04	119
32	2012-08-05	207
33	2012-08-06	201
34	2012-08-07	NaN
35	2012-08-08	NaN
36	2012-08-09	NaN
37	2012-08-10	229

Analytics Vidhya

Data Preprocessing in Time Series

S. No.	Date	Count
31	2012-08-04	119
32	2012-08-05	207
33	2012-08-06	201
34	2012-08-07	NaN
35	2012-08-08	NaN
36	2012-08-09	NaN
37	2012-08-10	229

208

215

222

Analytics Vidhya

Data Preprocessing in Time Series

S. No.	Date	Count
31	2012-08-04	119
32	2012-08-05	207
33	2012-08-06	201
34	2012-08-07	NaN
35	2012-08-08	NaN
36	2012-08-09	NaN
37	2012-08-10	229

208

215

222

- Interpolation is a commonly used method for imputing missing values.
- It creates new data points within the range of individual observations.
- We can use linear, polynomial or piecewise spline functions

Analytics Vidhya

Notebook

Analytics Vidhya

Thank You

Dealing with missing values in Time Series

- Mathematical equation ? (mention atleast linear here)
-

  Fits a line/curve through points. .. that’s all
  Out of context

Analytics Vidhya