

Seasonal dummy variables

Seasonality
features

Seasonal dummy variables

- Seasonal dummy variables are the features created by one hot encoding the date or time feature.

Date	Day of week	is_mon	is_tue	is_wed	is_thu	is_fri	is_sat
2023-05-08	0 (Mon)	1	0	0	0	0	0
2023-05-09	1 (Tue)	0	1	0	0	0	0
2023-05-10	2 (Wed)	0	0	1	0	0	0
2023-05-11	3 (Thu)	0	0	0	1	0	0
2023-05-12	4 (Fri)	0	0	0	0	1	0
2023-05-13	5 (Sat)	0	0	0	0	0	1
2023-05-14	6 (Sun)	0	0	0	0	0	0


Seasonal dummy variables

- To avoid the dummy variable trap we create one fewer feature than the number of categories.

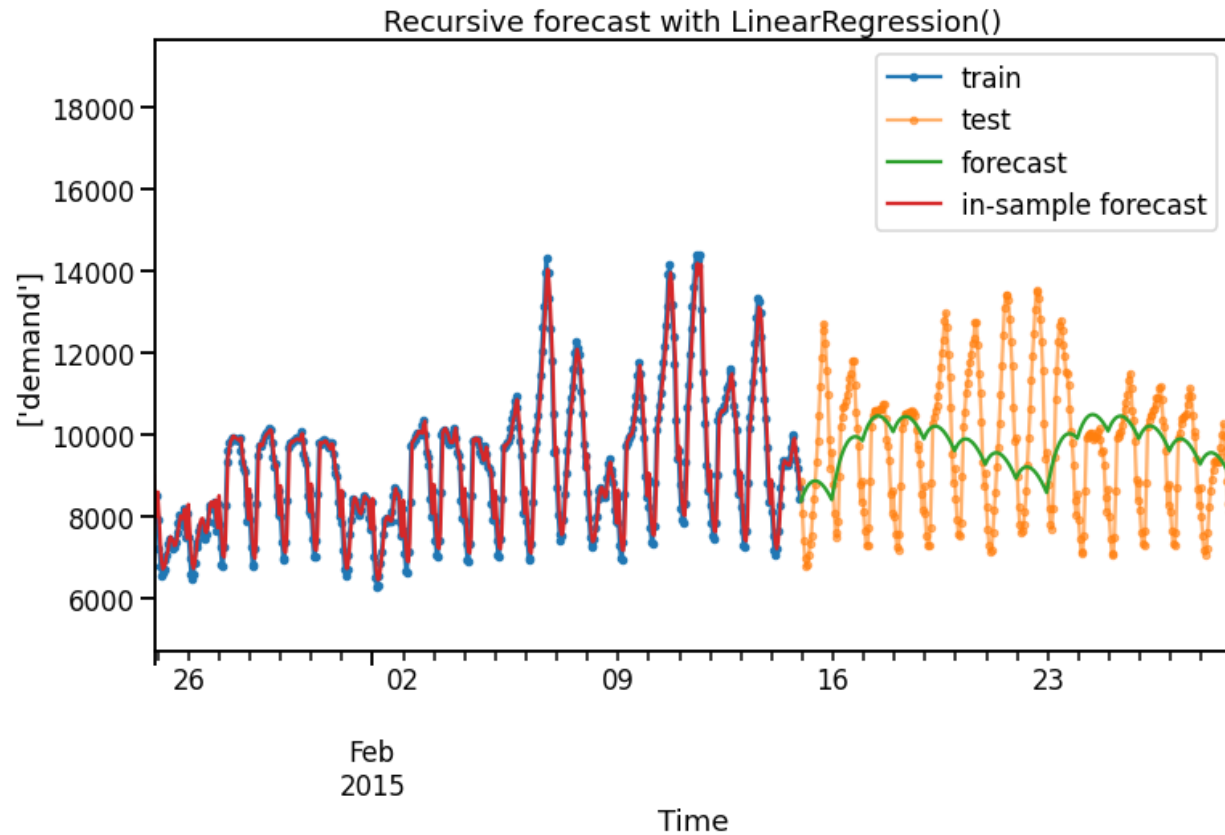
Date	Day of week	is_mon	is_tue	is_wed	is_thu	is_fri	is_sat
2023-05-08	0 (Mon)	1	0	0	0	0	0
2023-05-09	1 (Tue)	0	1	0	0	0	0
2023-05-10	2 (Wed)	0	0	1	0	0	0
2023-05-11	3 (Thu)	0	0	0	1	0	0
2023-05-12	4 (Fri)	0	0	0	0	1	0
2023-05-13	5 (Sat)	0	0	0	0	0	1
2023-05-14	6 (Sun)	0	0	0	0	0	0

When are seasonal dummies helpful?

- Seasonal dummy variables are useful when working with linear models.
- It allows linear models to estimate the impact of each seasonal value (e.g., Monday) separately.
- This overcomes the limitations of using the numeric date and time features with linear models.
- Downside: they can create a lot of additional features (e.g., week of year, day of year).

$$y = \beta_0 + \beta_1 is_mon + \beta_2 is_tue + \beta_3 is_wed + \dots$$


Example: Electricity demand using datetime features



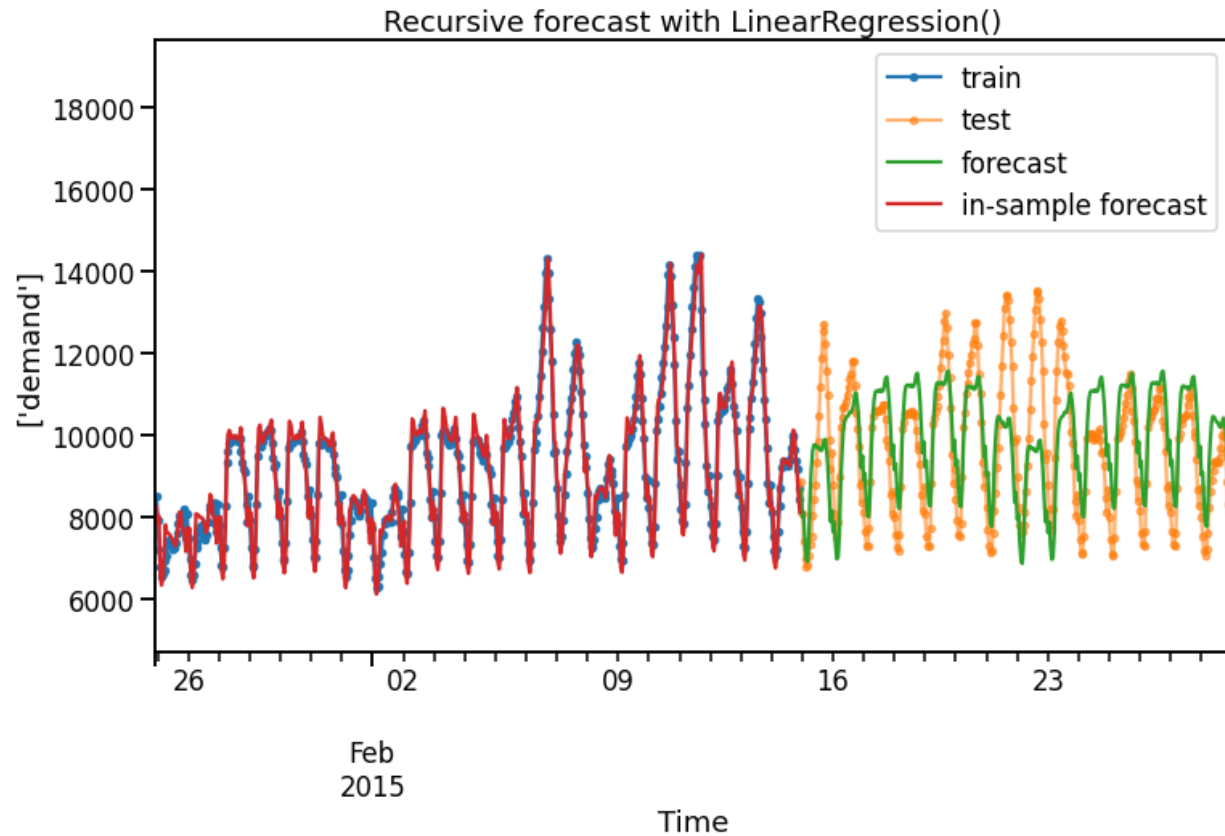
Features

- Trend feature: t ,
- Lag of 1 hour: y_{t-1}
- **Hour of day**
- **Day of week**
- **Month of the year**

Model

- Linear regression

Example: Electricity demand using seasonal dummies



Features

- Trend feature: t ,
- Lag of 1 hour: y_{t-1}
- **is_hour_1, is_hour_2, ...**
- **is_mon, is_tue, ...**
- **is_jan, is_feb, ...**

Model

- Linear regression

Summary

Seasonal dummy variables are one hot encodings of date and time features.

Enables linear models to estimate the impact of each seasonal value separately.

Downside is that we create a lot of additional features.