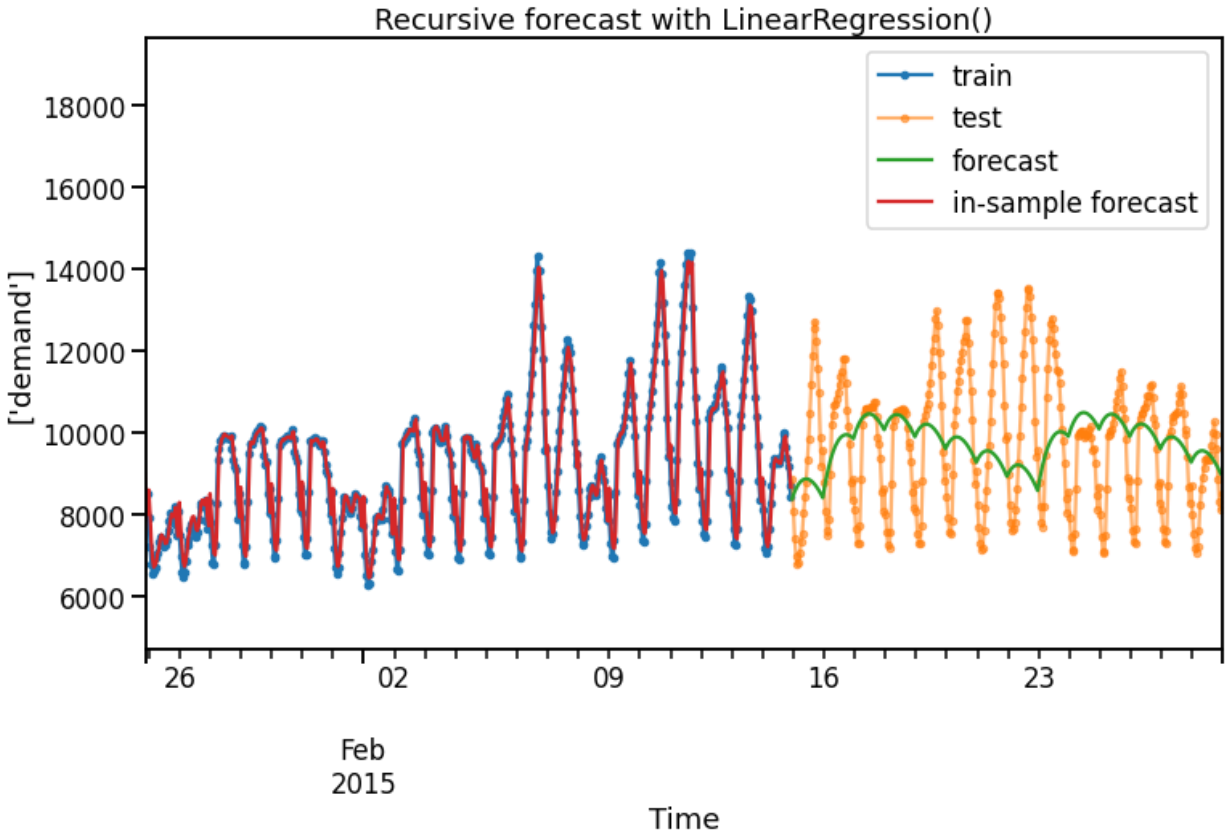
Why do linear models struggle with datetime features?

Seasonality features

Example: Electricity demand



Features

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

Model

- Linear regression

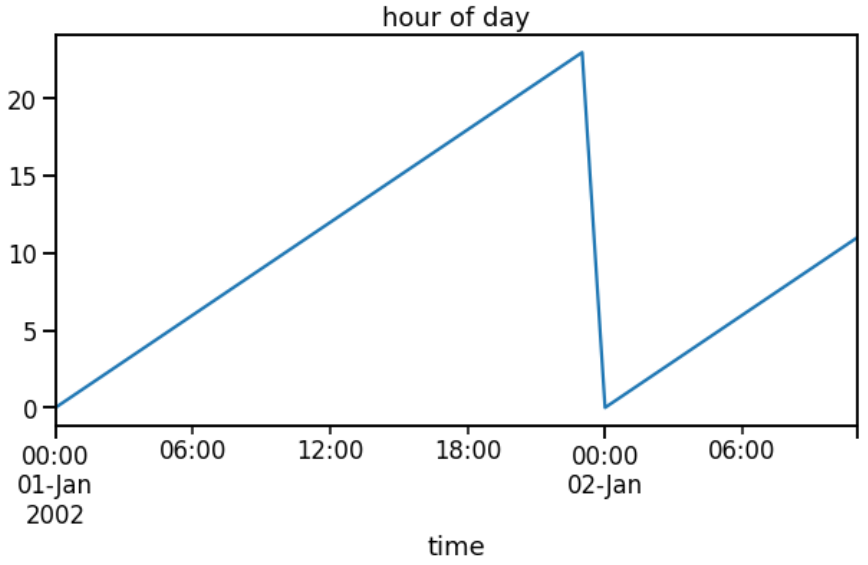
Why do linear models struggle with datetime features?

- When computing these features in most packages we receive numeric features.

Hour of day	Day of week	Month of year
1	2	2
2	2	2
3	2	2
4	2	2

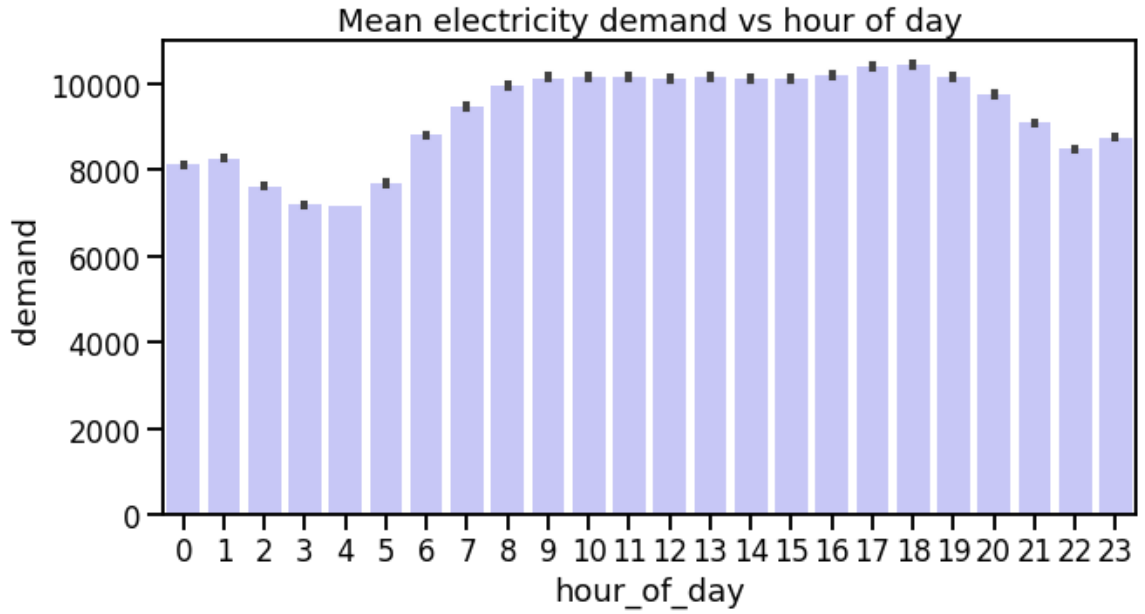
Why do linear models struggle with datetime features?

- When computing these features in most packages we receive numeric features.
- Most of these variables are cyclical. The numeric representation does not capture this.



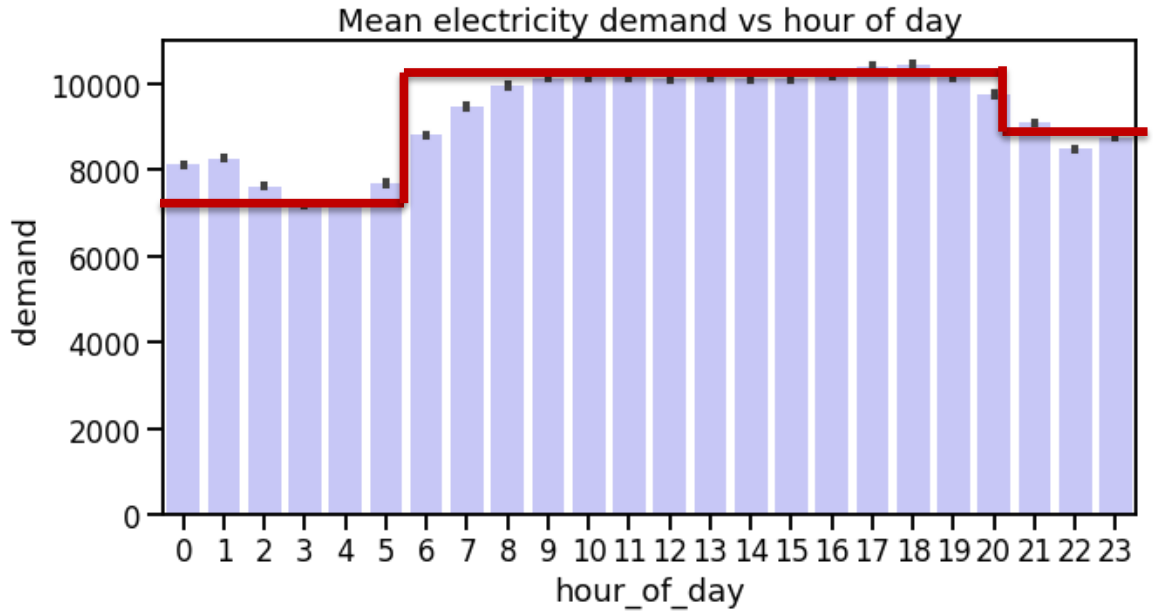
Why do linear models struggle with datetime features?

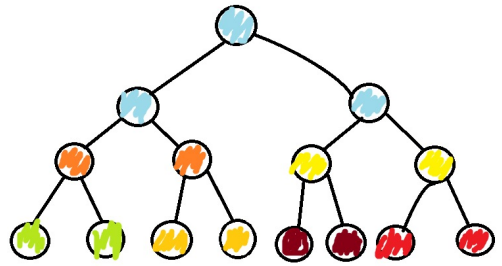
- When computing these features in most packages we receive numeric features.
- Most of these variables are cyclical. The numeric representation does not capture this.
- Cyclical variables often have non-linear relationships with the target.



Why do linear models struggle with datetime features?

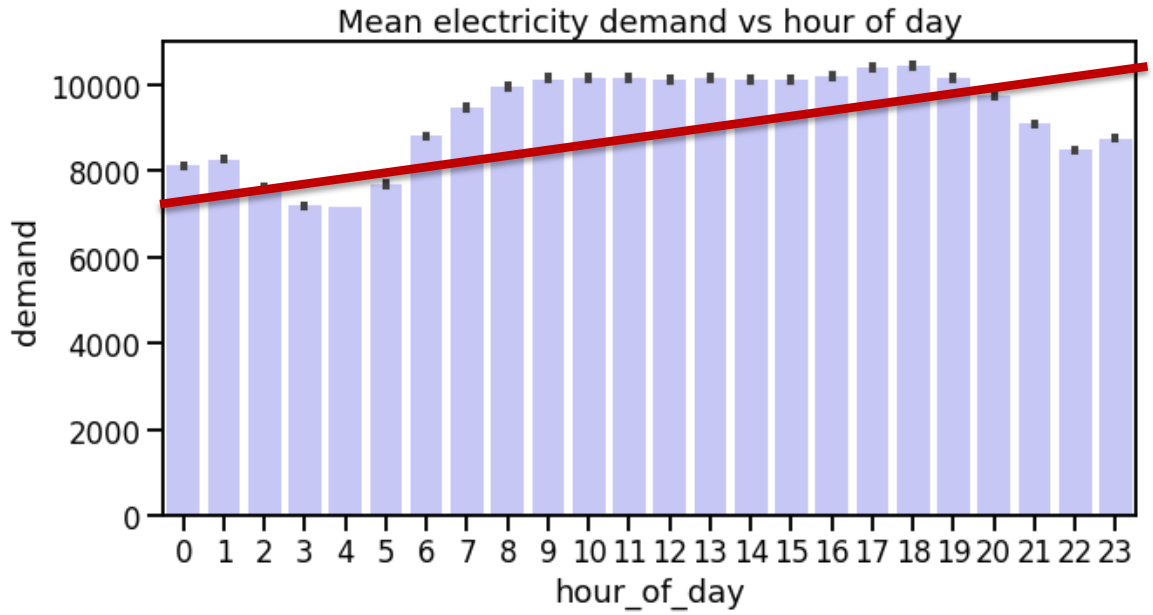
- When computing these features in most packages we receive numeric features.
- Most of these variables are cyclical. The numeric representation does not capture this.
- Cyclical variables often have non-linear relationships with the target.
- Tree-based models can model the non-linear relationship between the target and features.





Why do linear models struggle with datetime features?

- When computing these features in most packages we receive numeric features.
- Most of these variables are cyclical. The numeric representation does not capture this.
- Cyclical variables often have non-linear relationships with the target.
- Tree-based models can model the non-linear relationship between the target and features.
- Linear models are constrained to fit a linear relationship between the target and features.

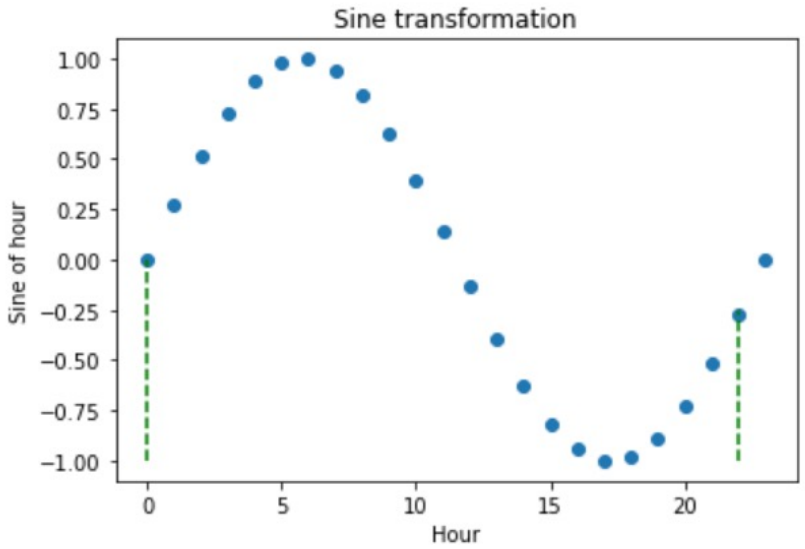


$$y = \beta_0 + \beta_1 hour\_of\_day + \beta_2 x_2 + \cdots$$

Why do linear models struggle with datetime features?

- When computing these features in most packages we receive numeric features.
- Most of these variables are cyclical. The numeric representation does not capture this.
- Cyclical variables often have non-linear relationships with the target.
- Tree-based models can model the non-linear relationship between the target and features.
- Linear models are constrained to fit a linear relationship between the target and features.
- Additional feature engineering can help linear models better use date & time variables!

Cyclical features:



Treat datetime features as a categorical variable and use:

- **One hot encoding (seasonal dummies)**
- **Target encoding**

Summary

Using features directly from the date and time can help capture multiple seasonalities.

Easy to compute and works well with tree-based models.

Does not work well with linear models.