Role of Feature Engineering in Machine Learning

Ravikiran Srinivasulu
SOFTWARE CONSULTANT

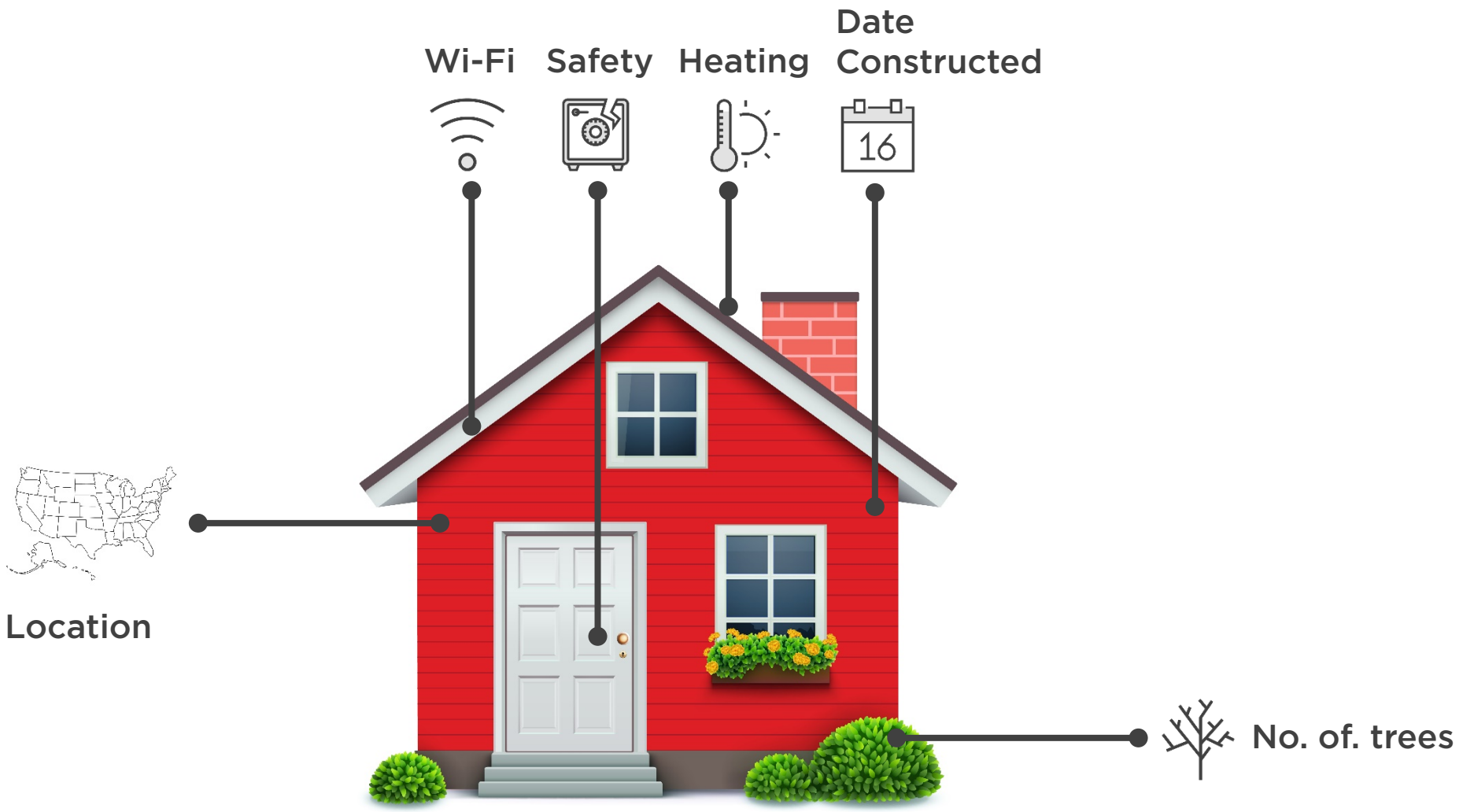
ravikirans.com | ravikirans.com/YouTube

Agenda

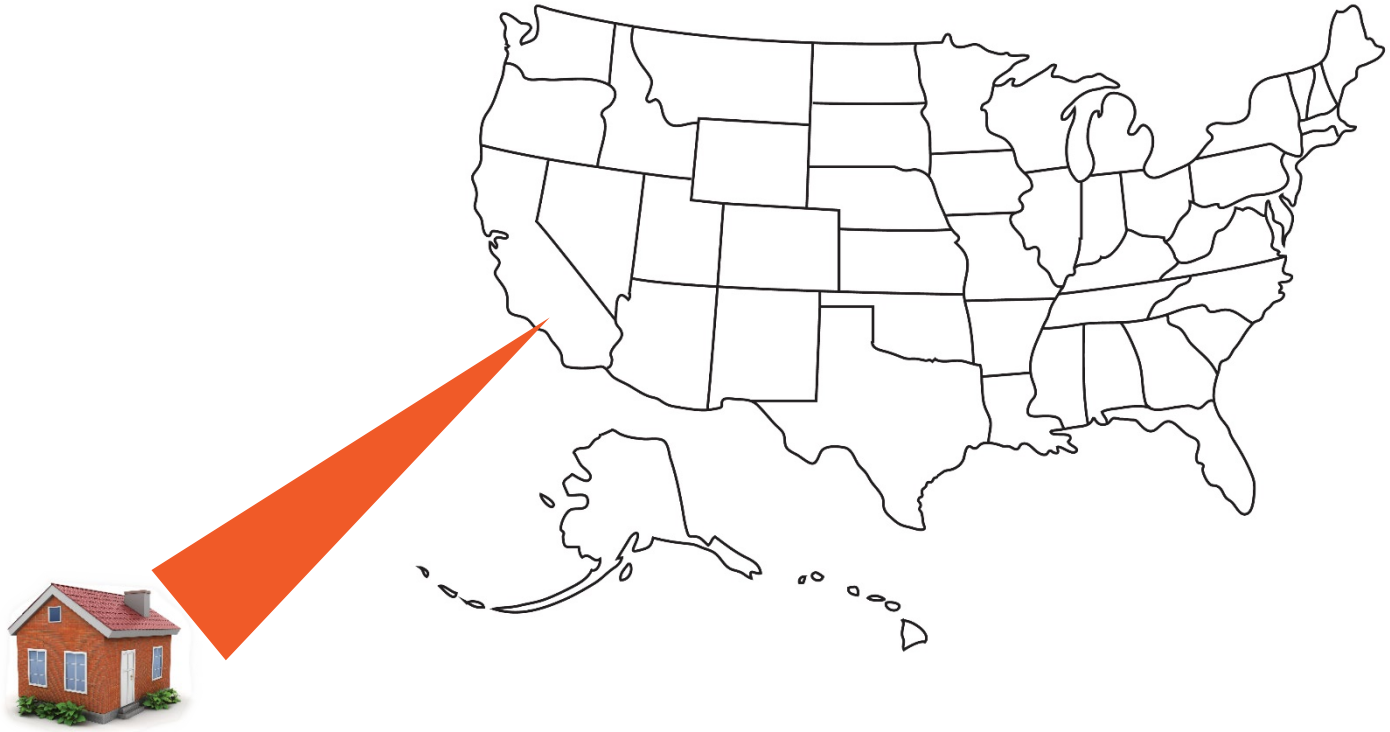
- Importance of Feature Engineering
- How Feature Engineering affect Model Complexity
- Use Feature Engineering to build better models
- Examples of Feature Engineering
- One-hot encoding & Learning by counts

Why Feature Engineering?

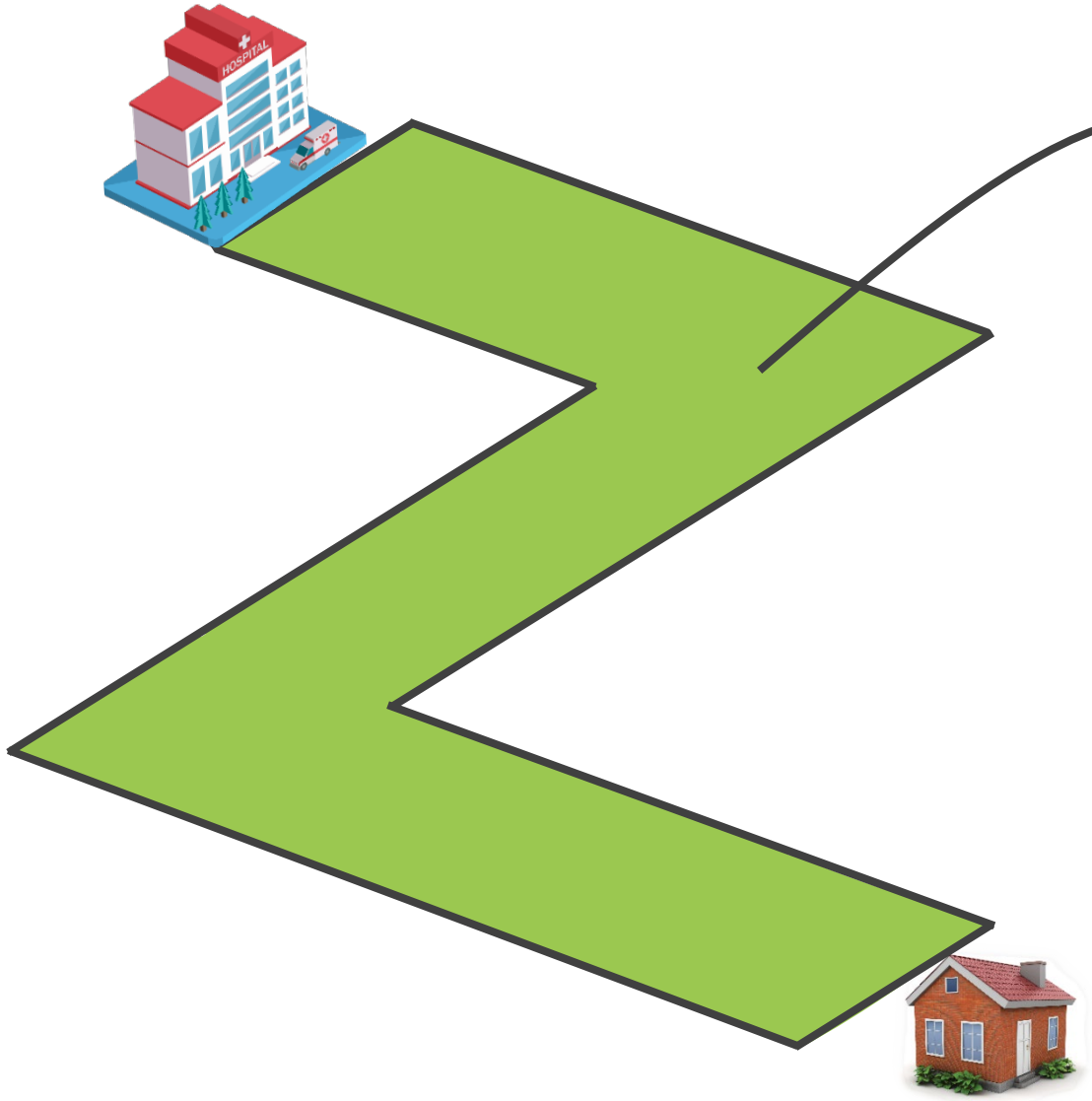
Predicting Rental Price of a House



33°97'75.2"N,
118°35'04.0"W

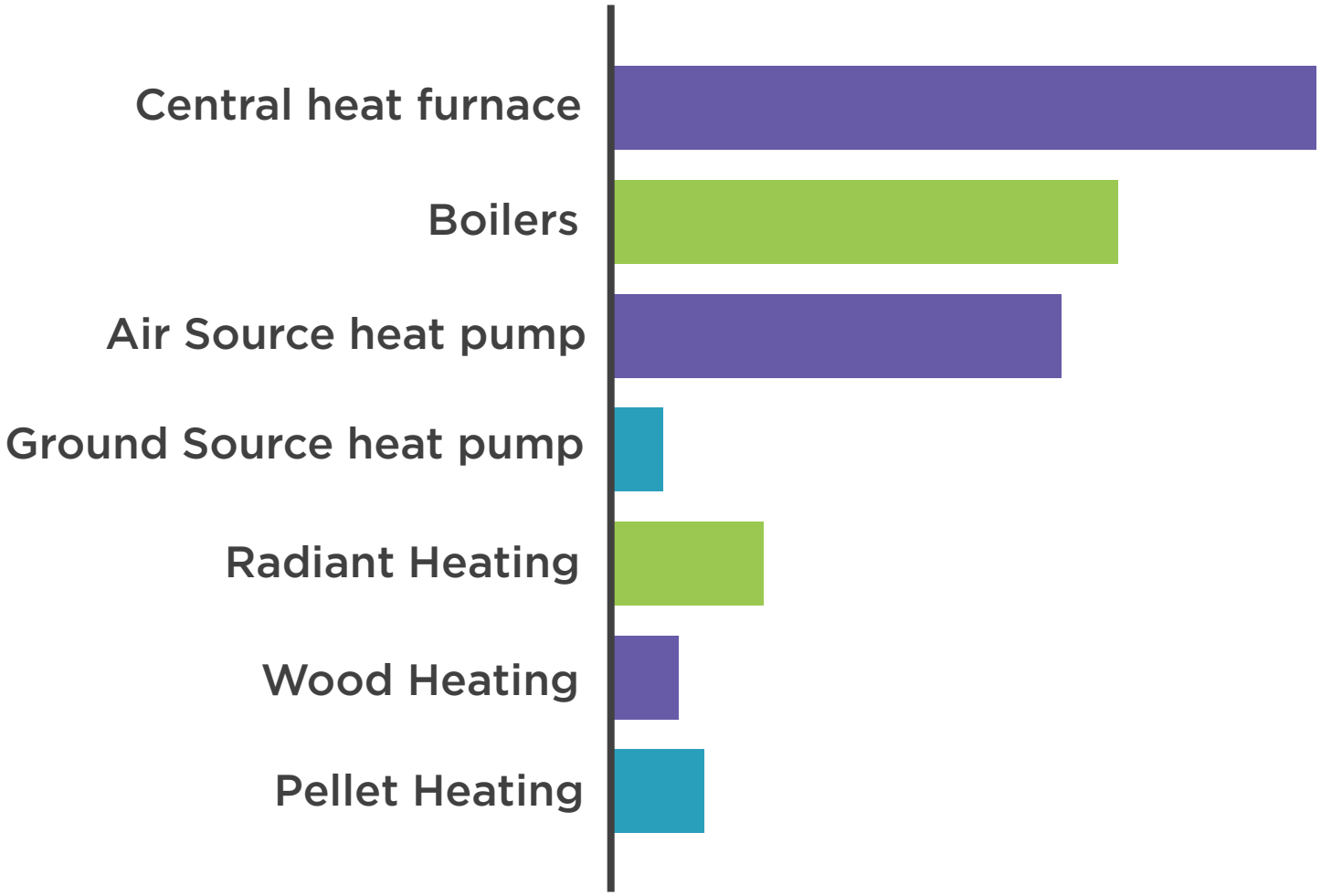


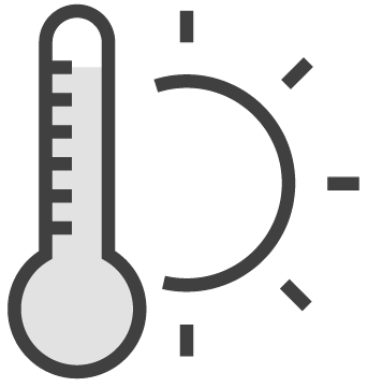
33°96'75.1"N,
118°30'49.2"W



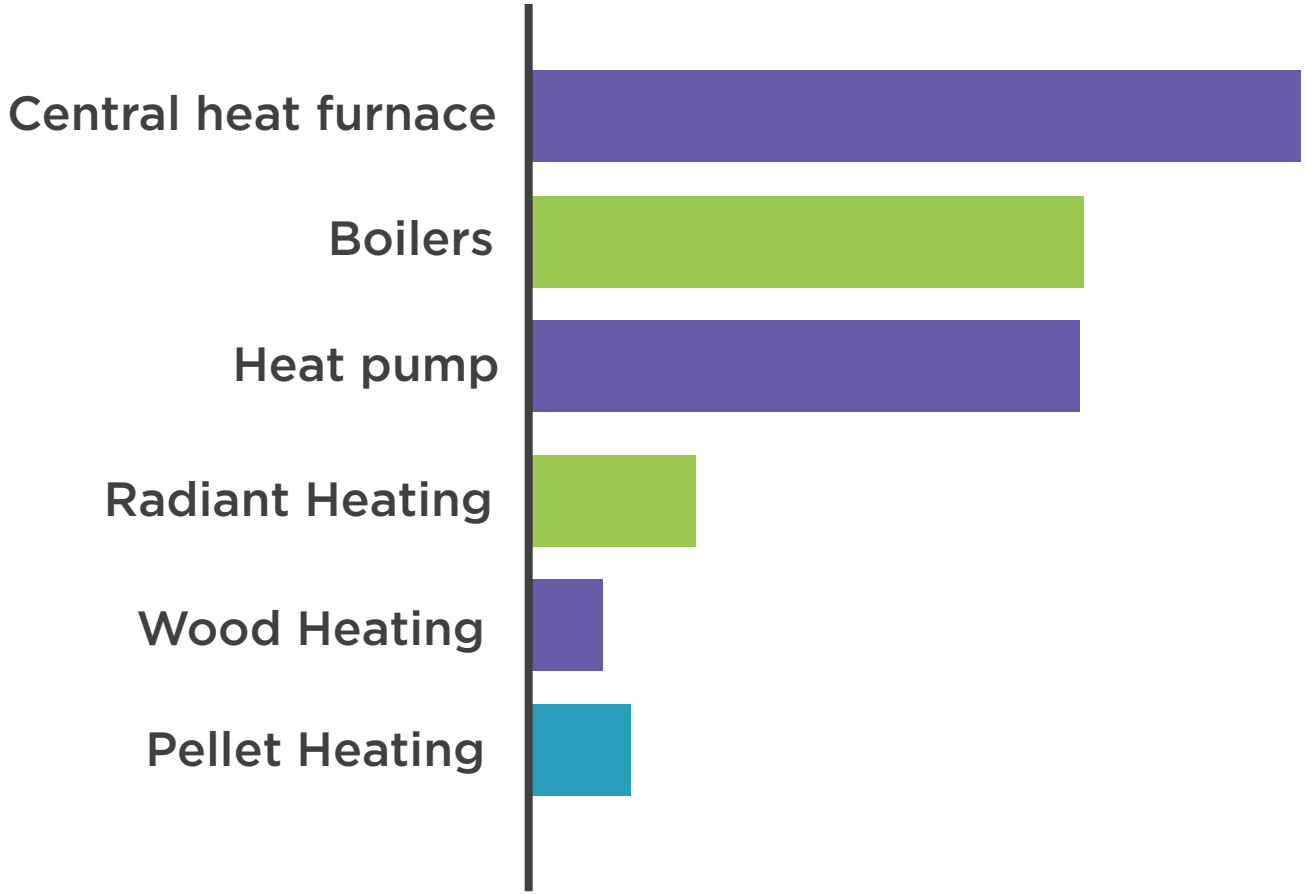
Use distance as a feature

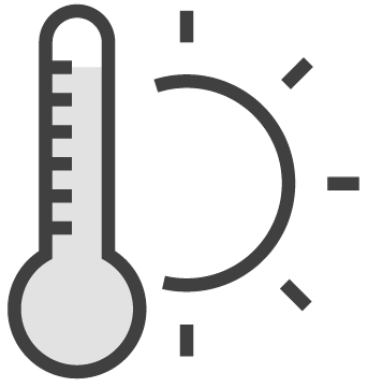
Predicting Rental Price of a House



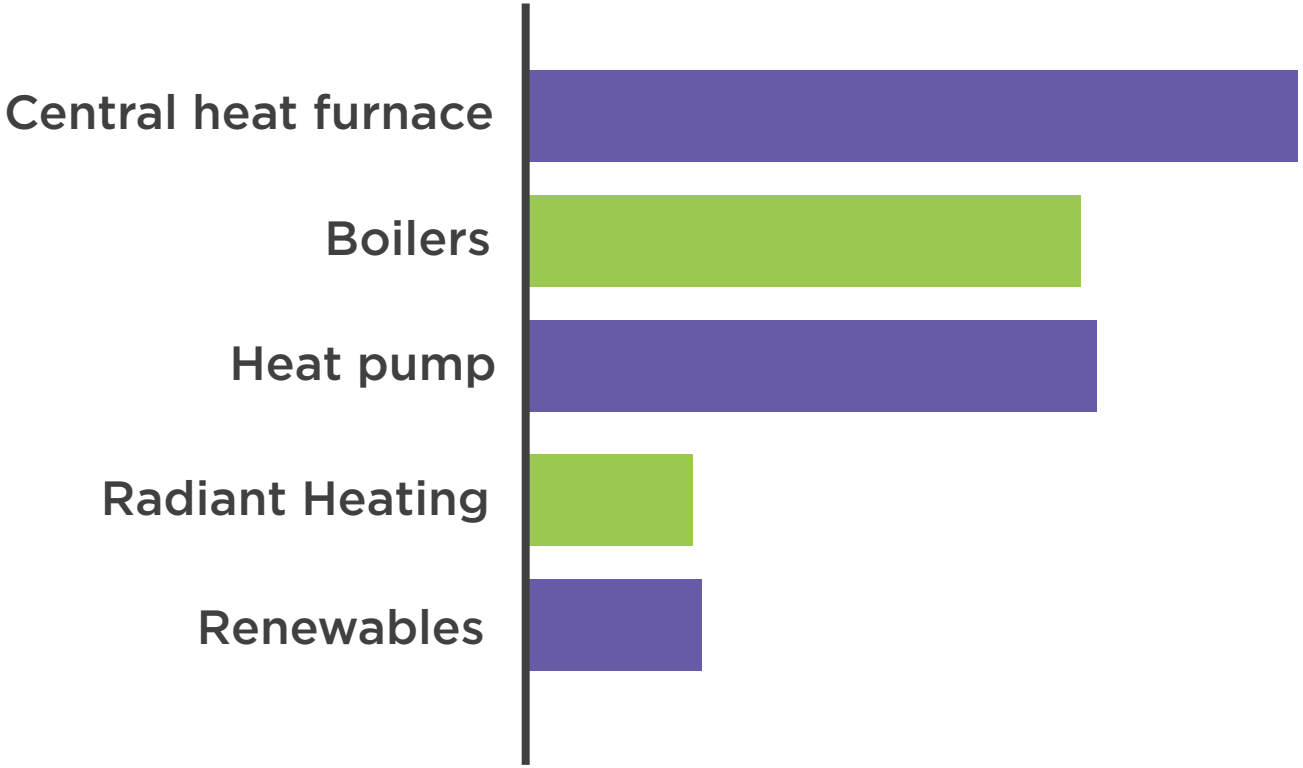


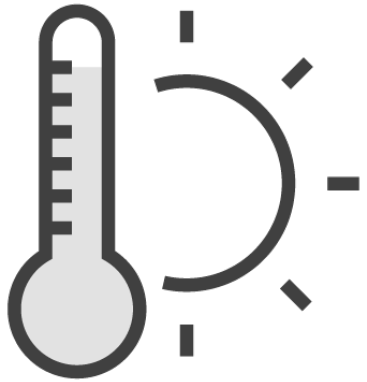
Predicting Rental Price of a House



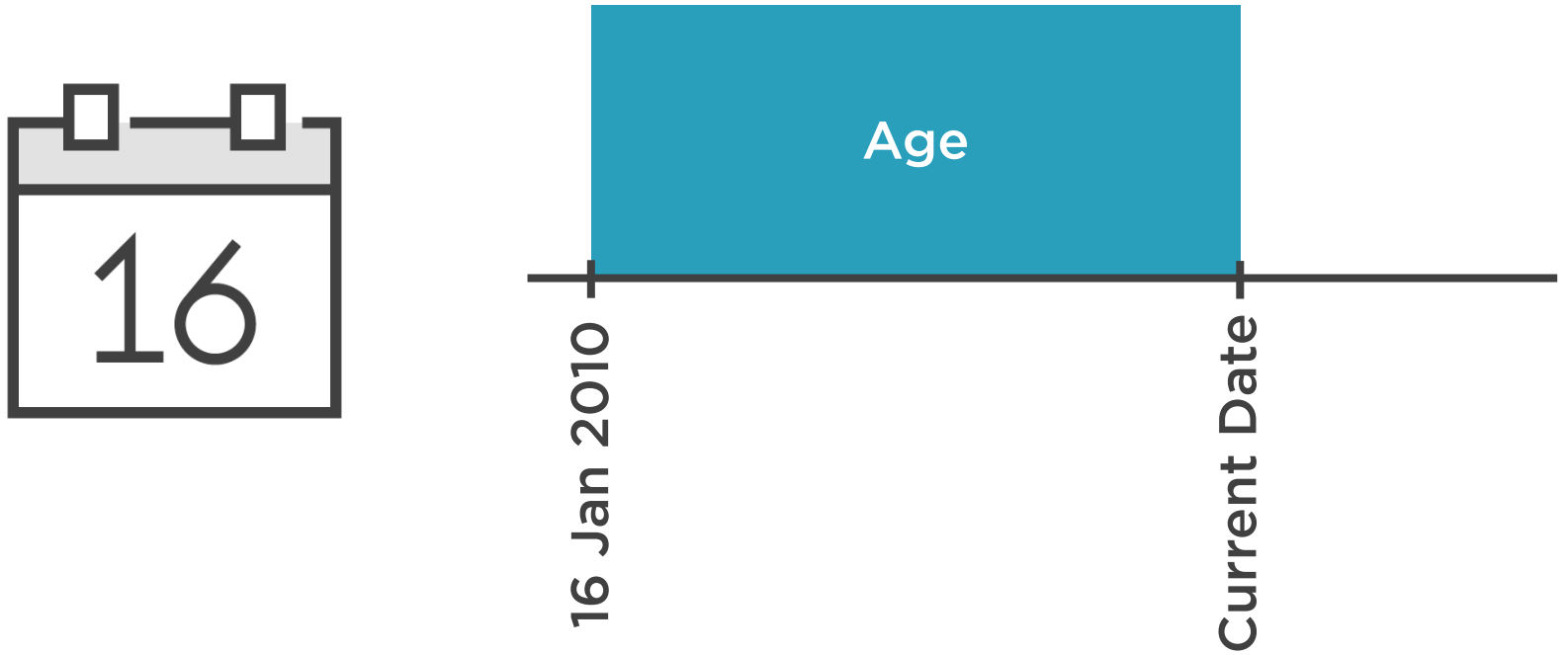


Predicting Rental Price of a House



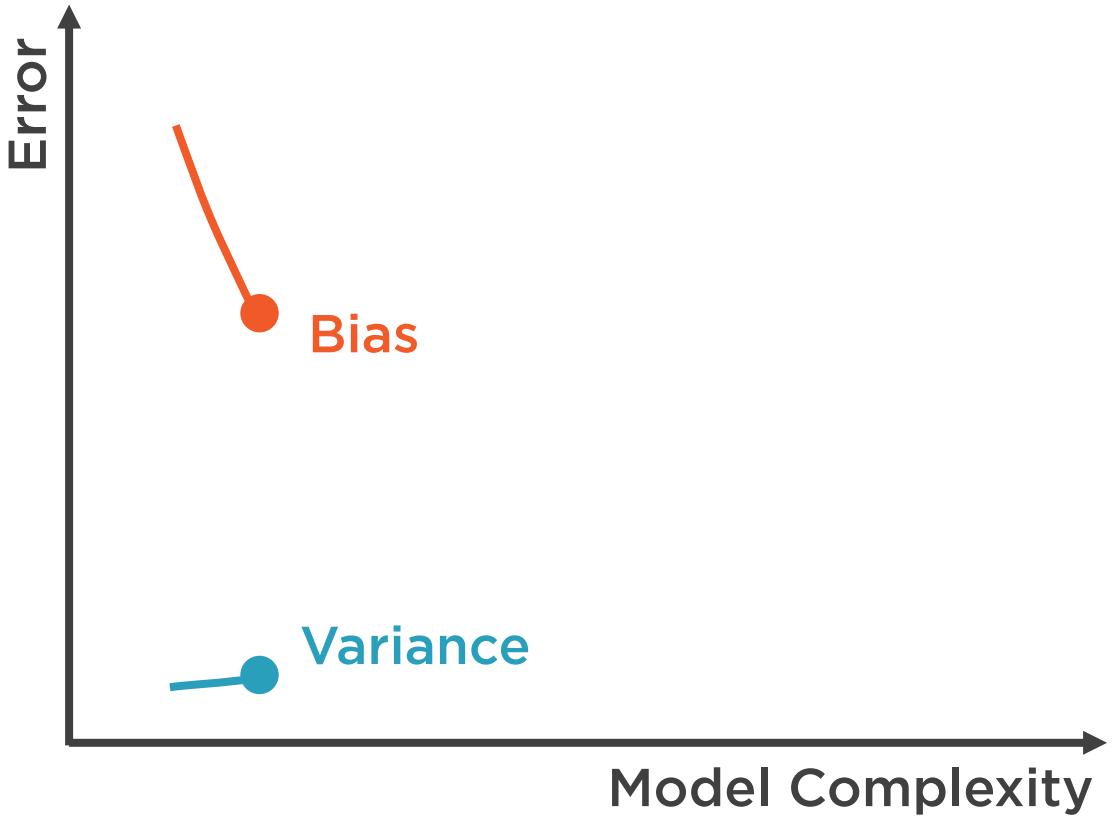


Predicting Rental Price of a House

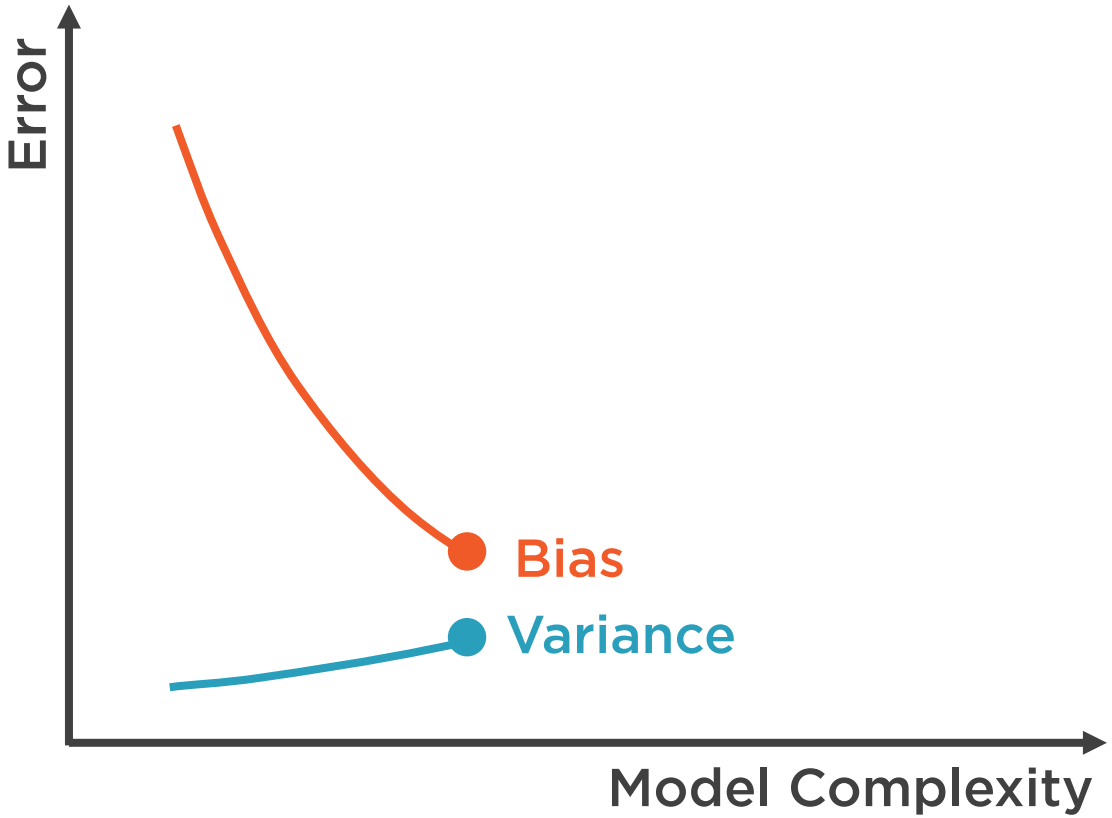


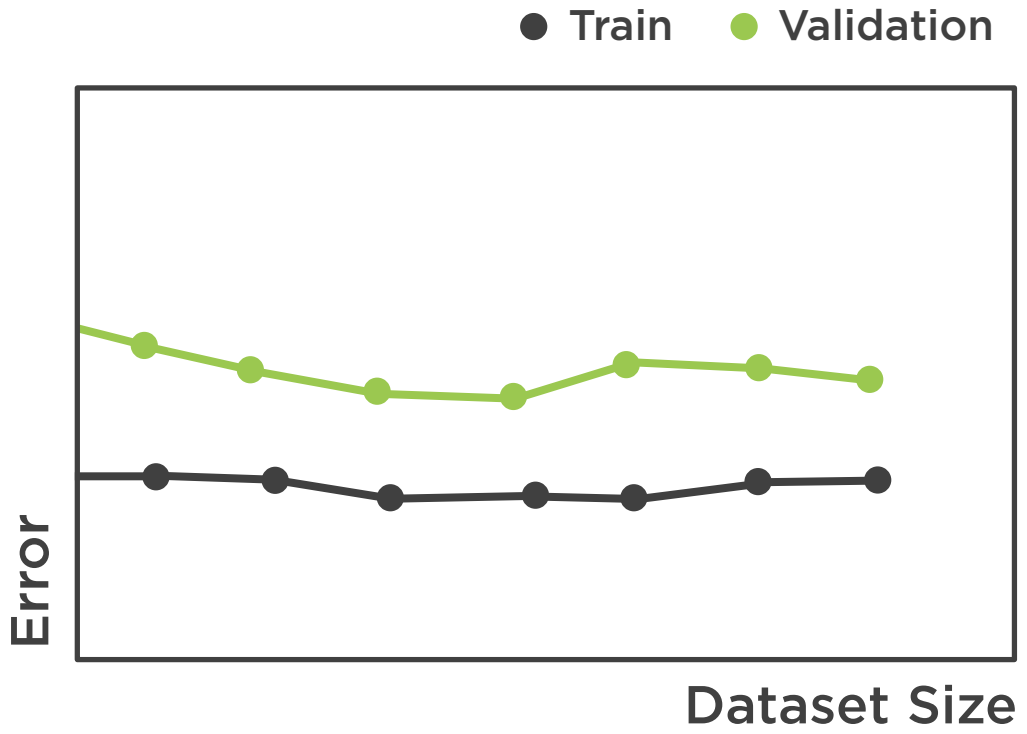
Role of Feature Engineering in Model Complexity

Model Complexity

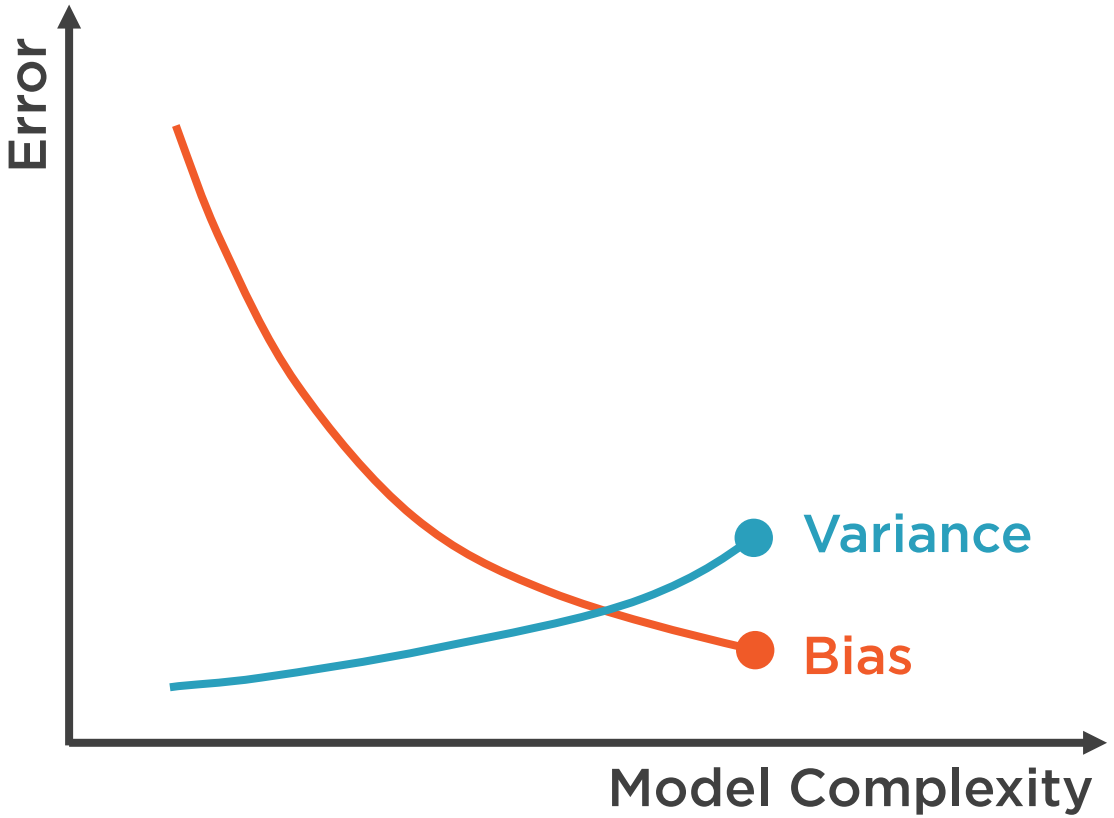


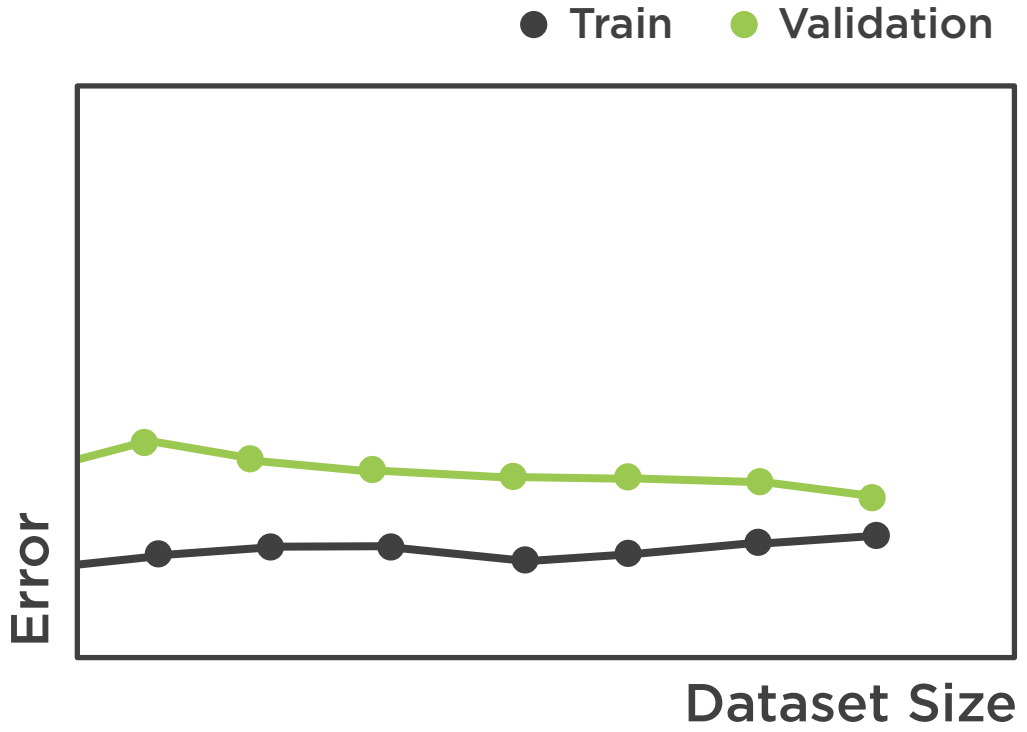
Model Complexity



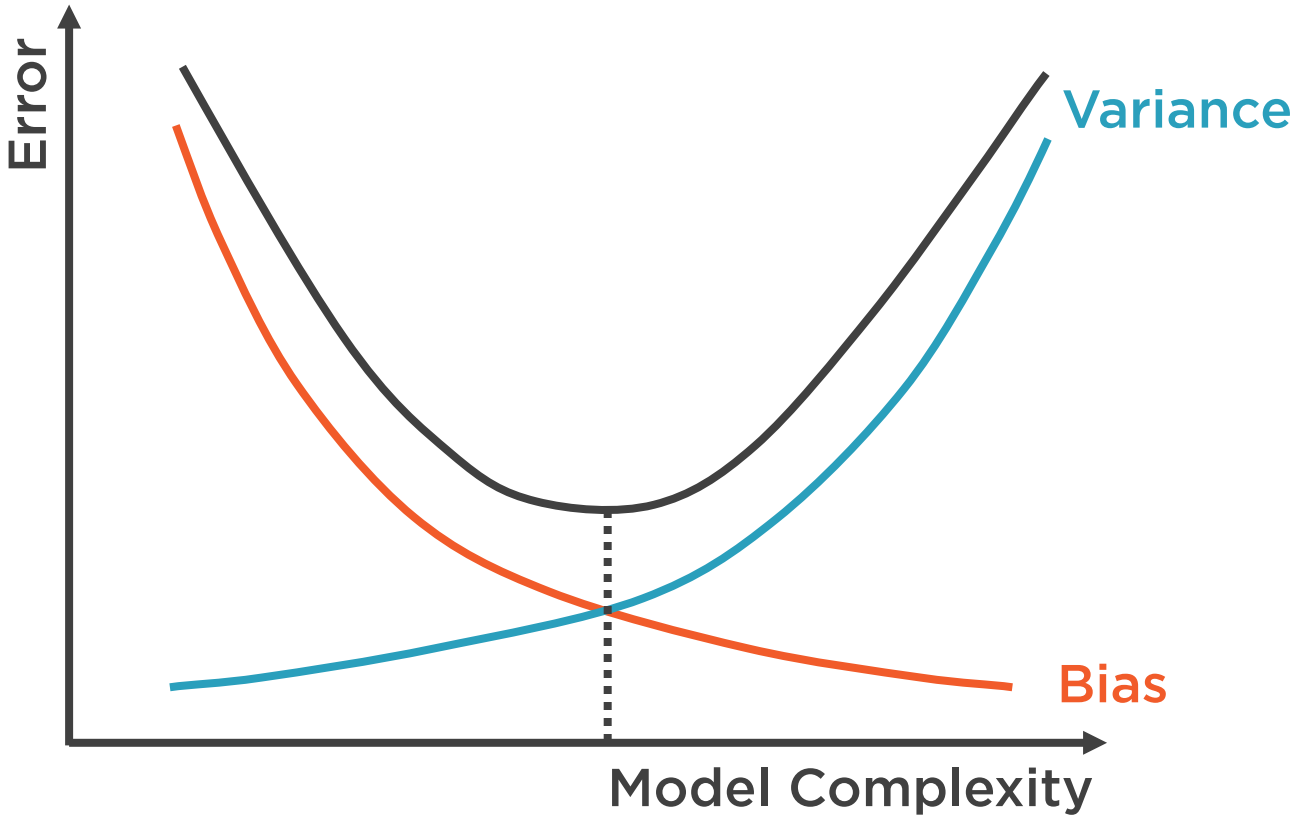


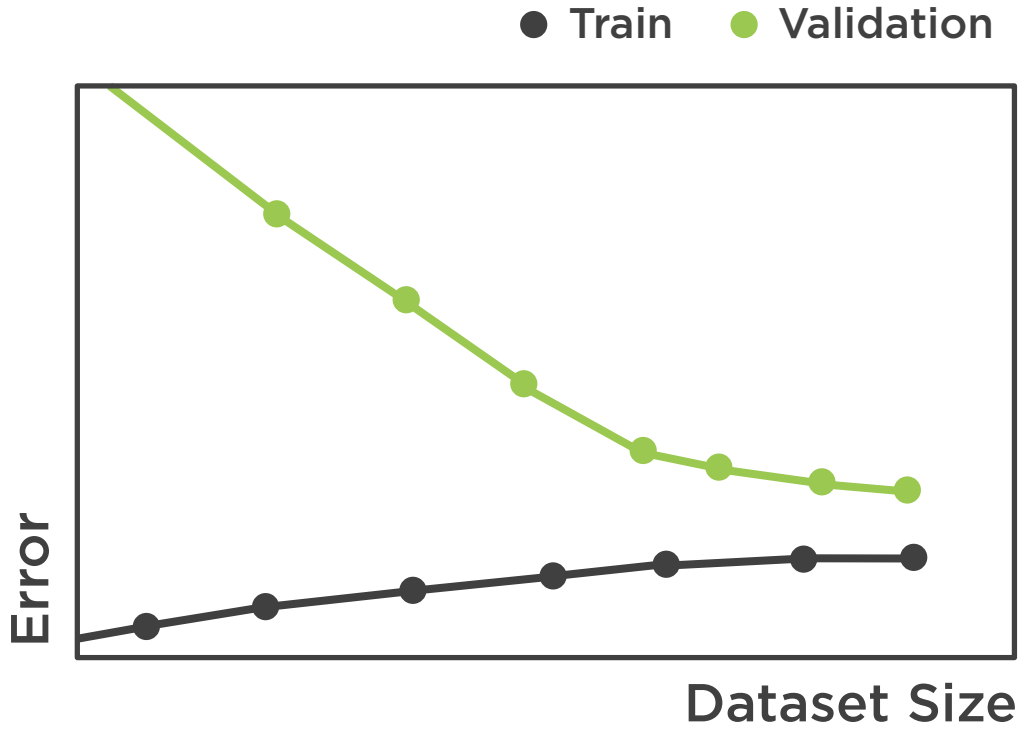
Model Complexity



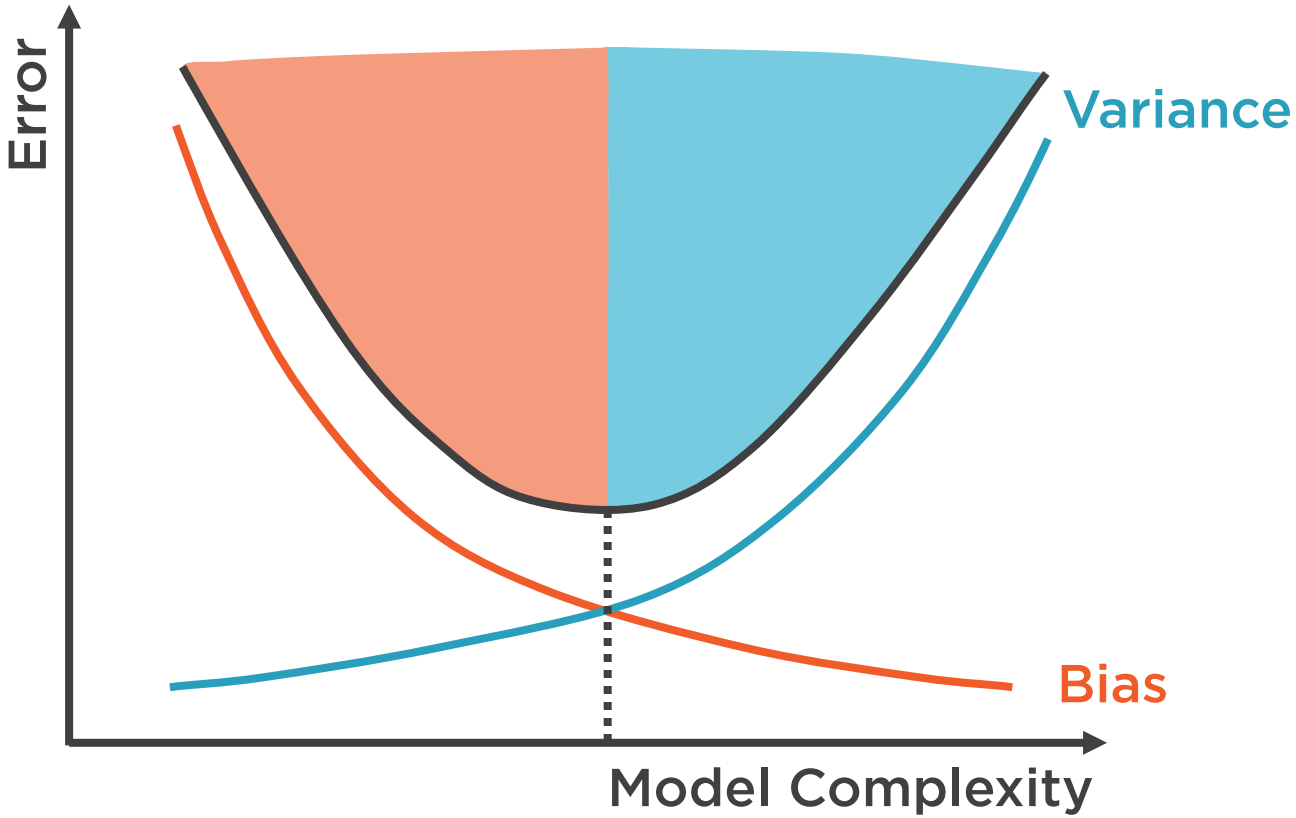


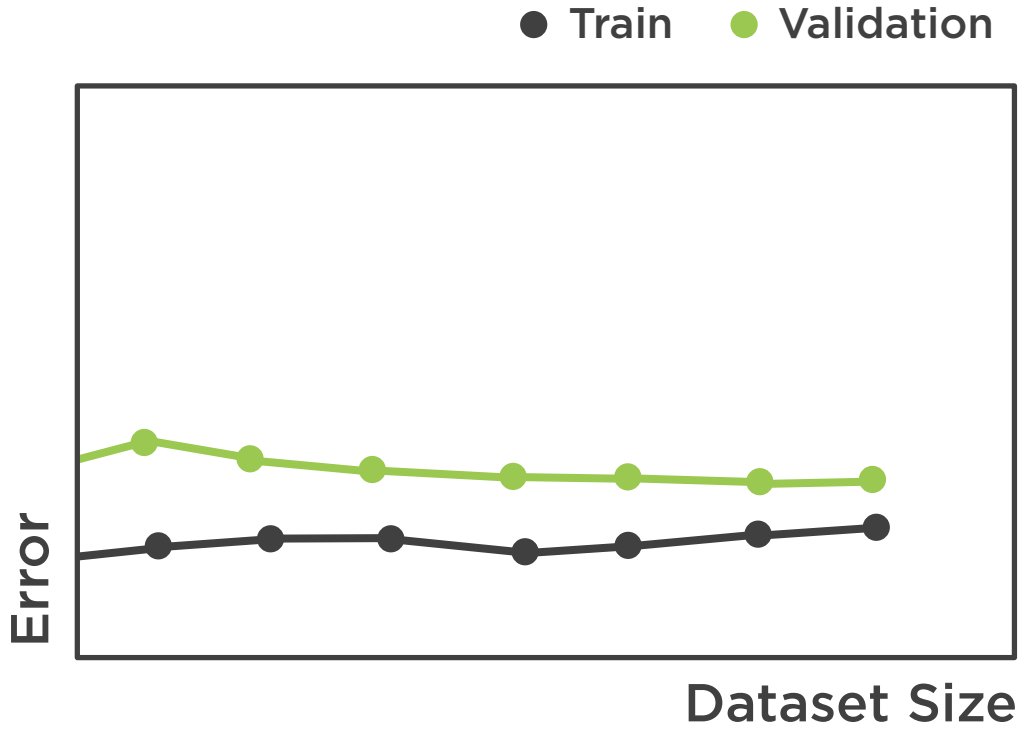
Model Complexity





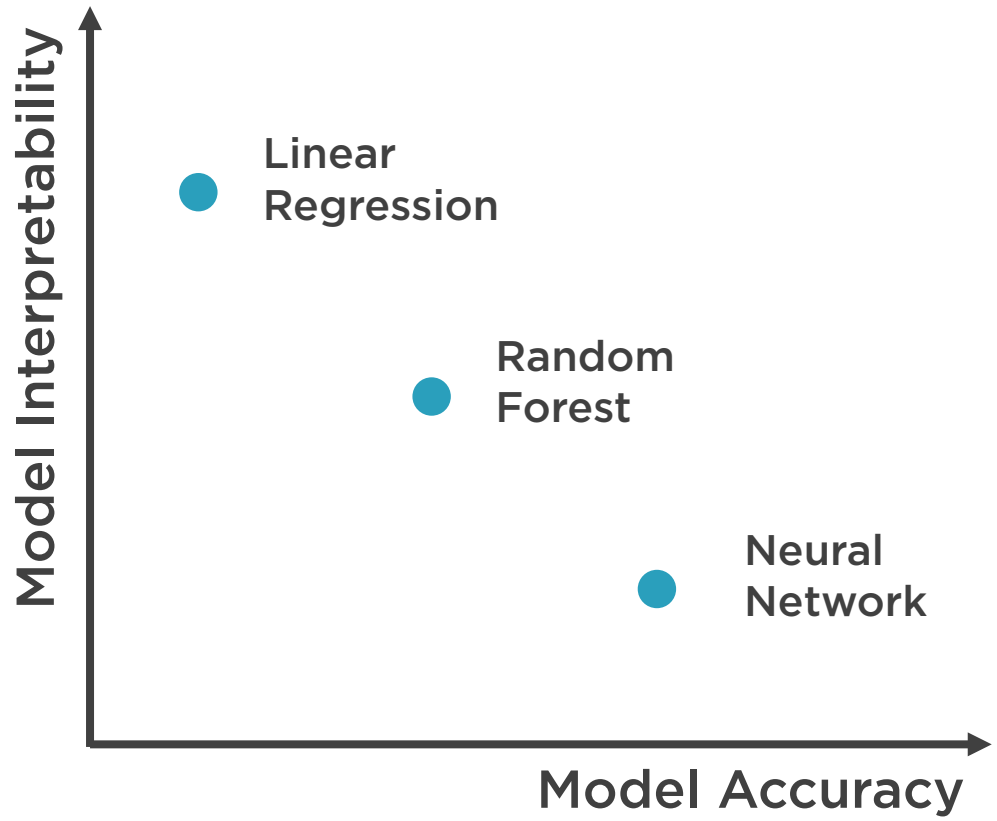
Model Complexity





Build Better Models With Feature Engineering

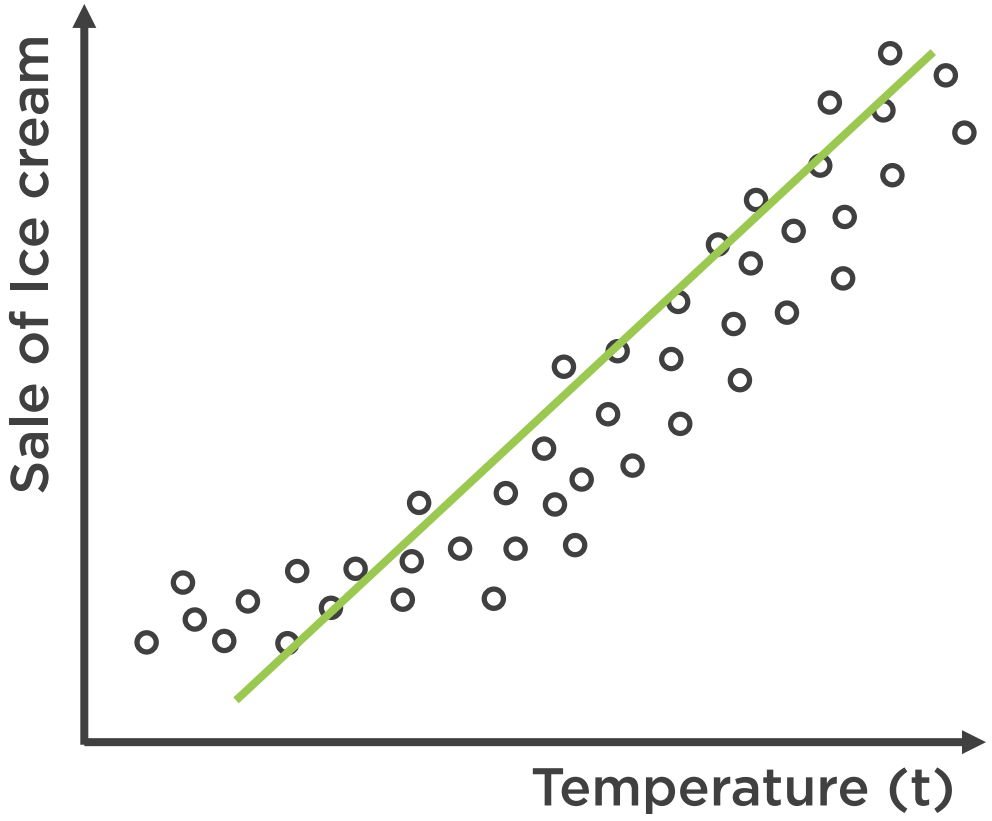
Model Accuracy vs. Model Interpretability

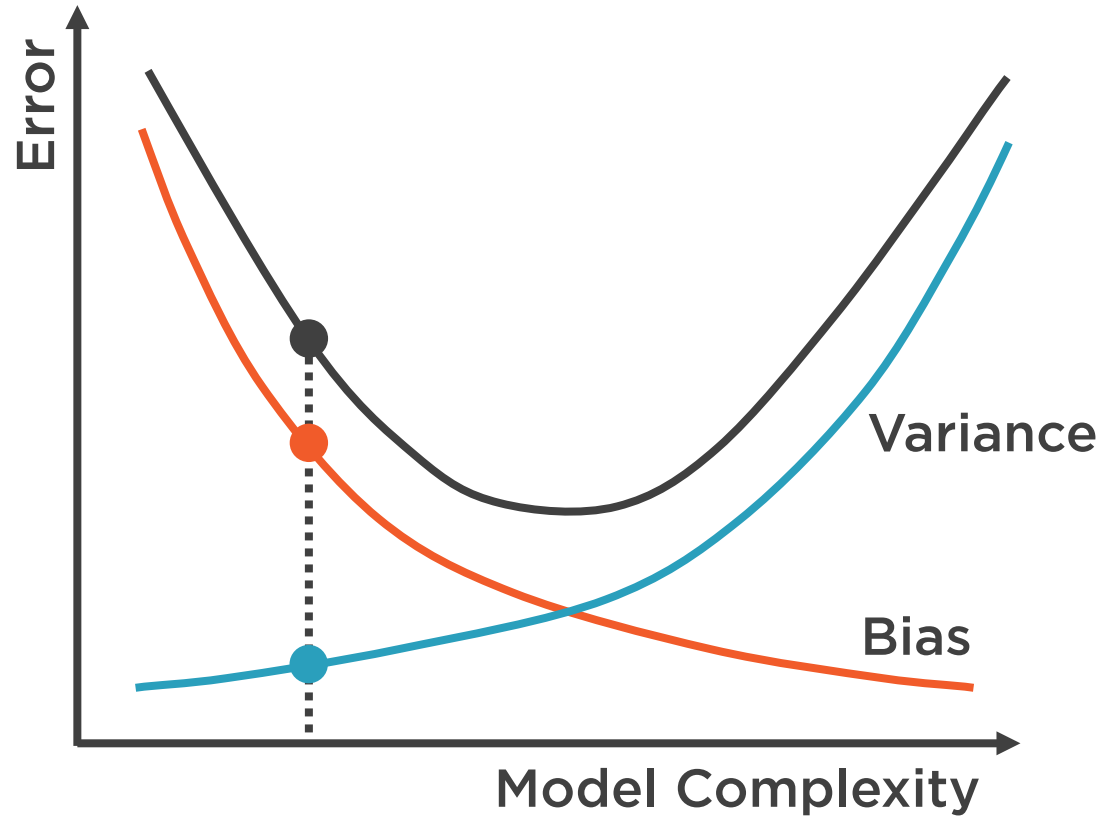


Linear Models:

- **Run faster**
- **Highly scalable**
- **Easier to train**

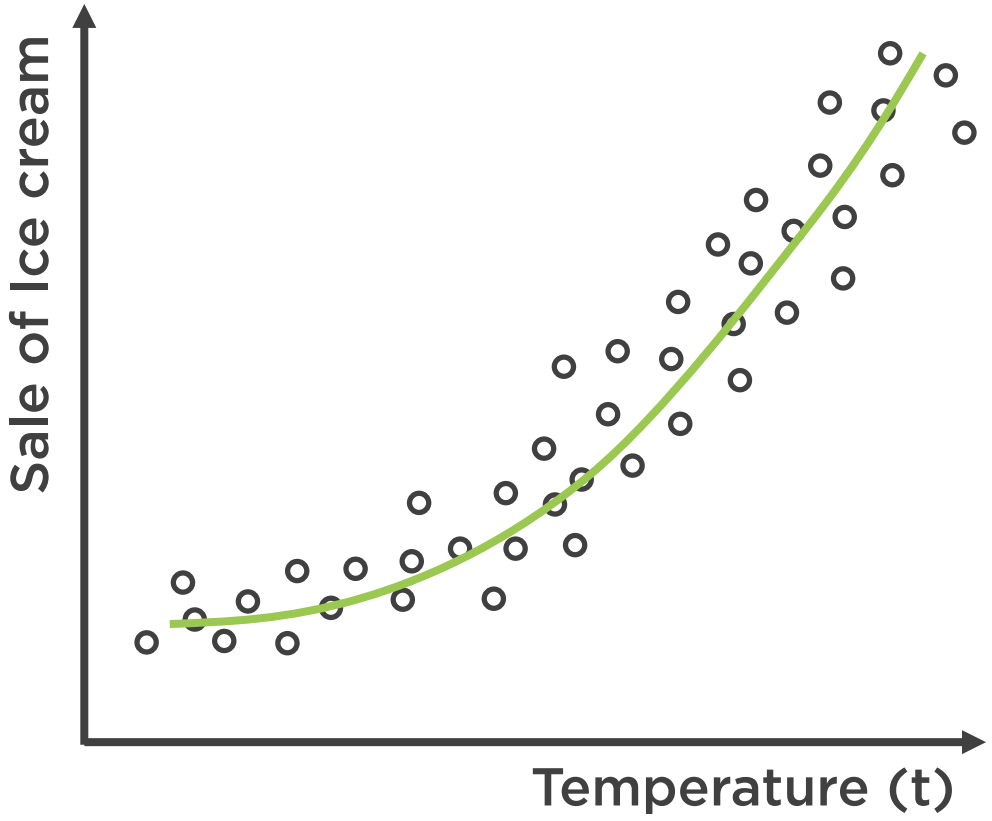
Building Linear Models

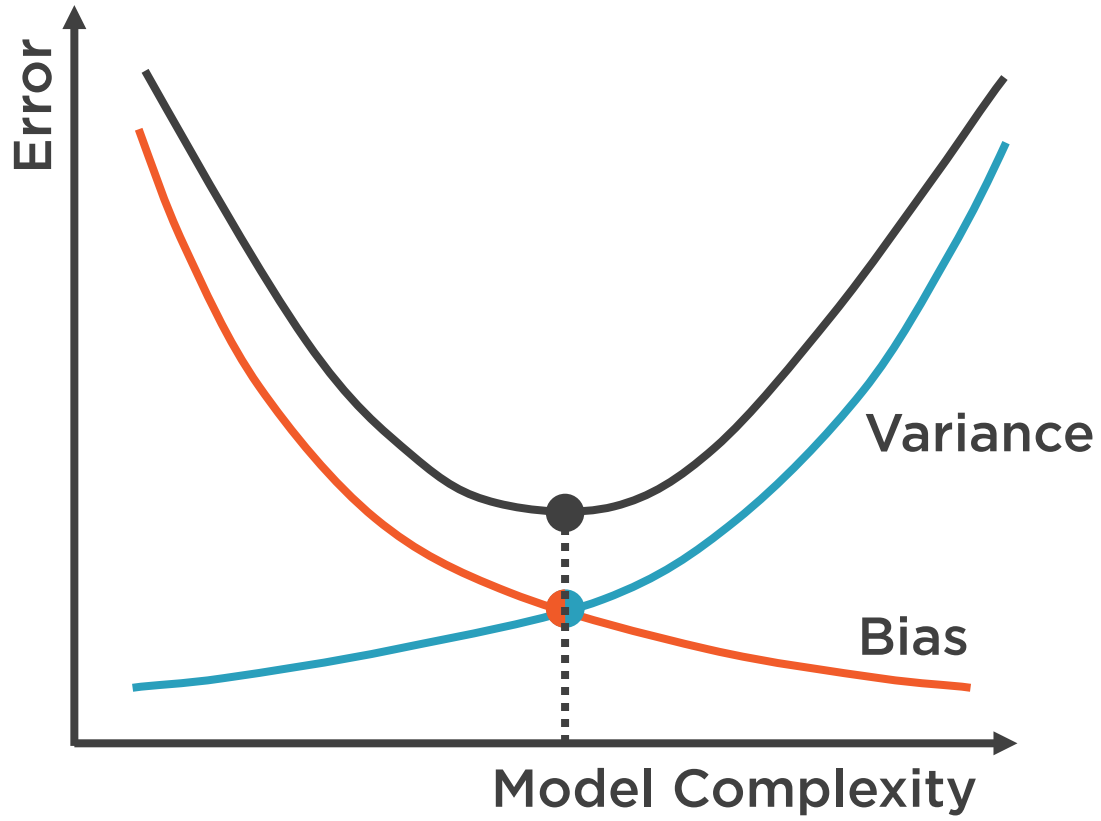




Sale = a_1(t) + C

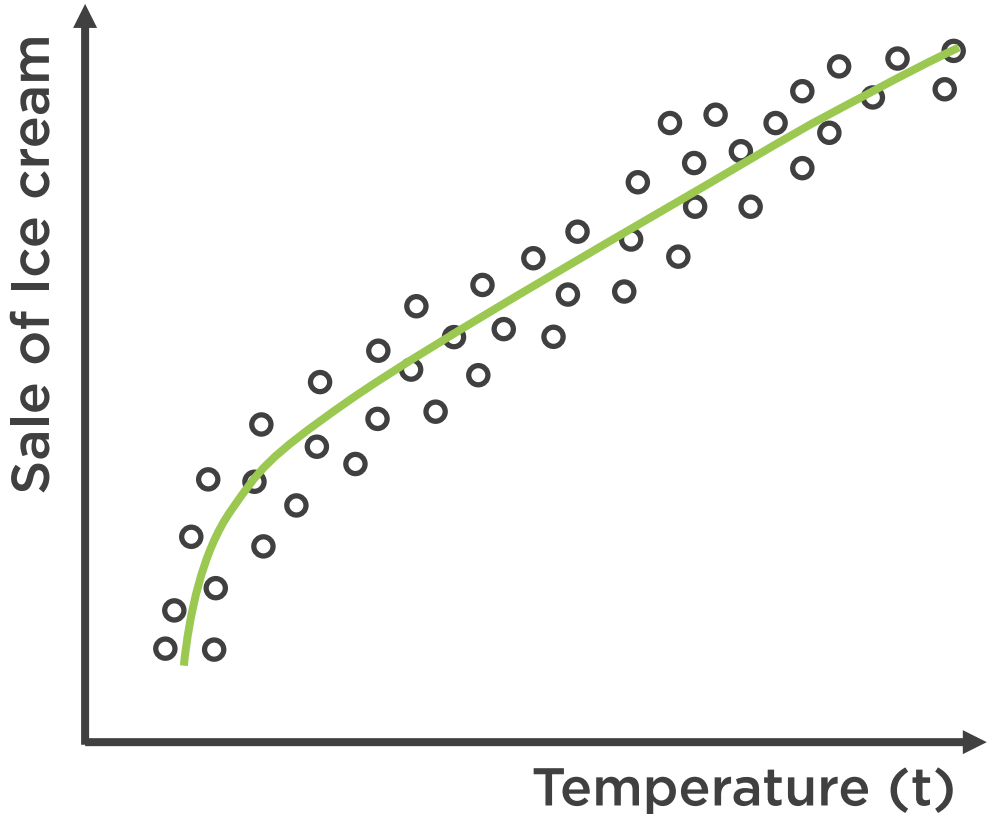
Building Linear Models





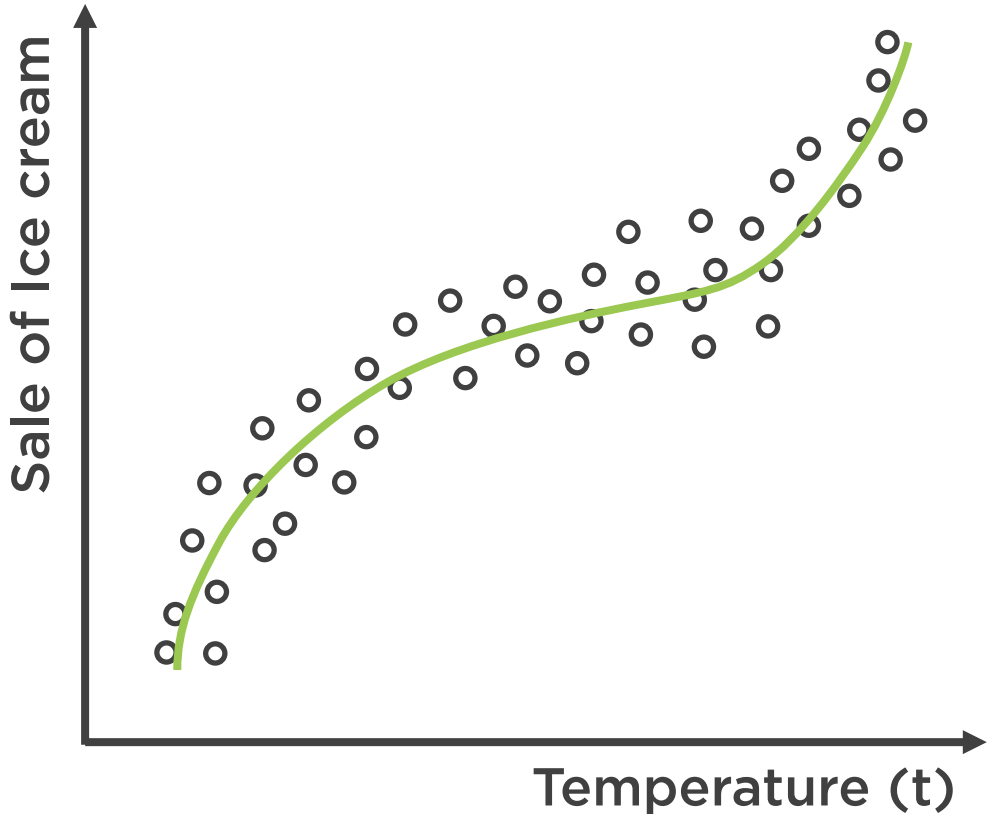
Sale = $a_1(t) + a_2(t)^2 + C$

Building Linear Models



Sale = a_1(t) + a_2(log(t)) + C

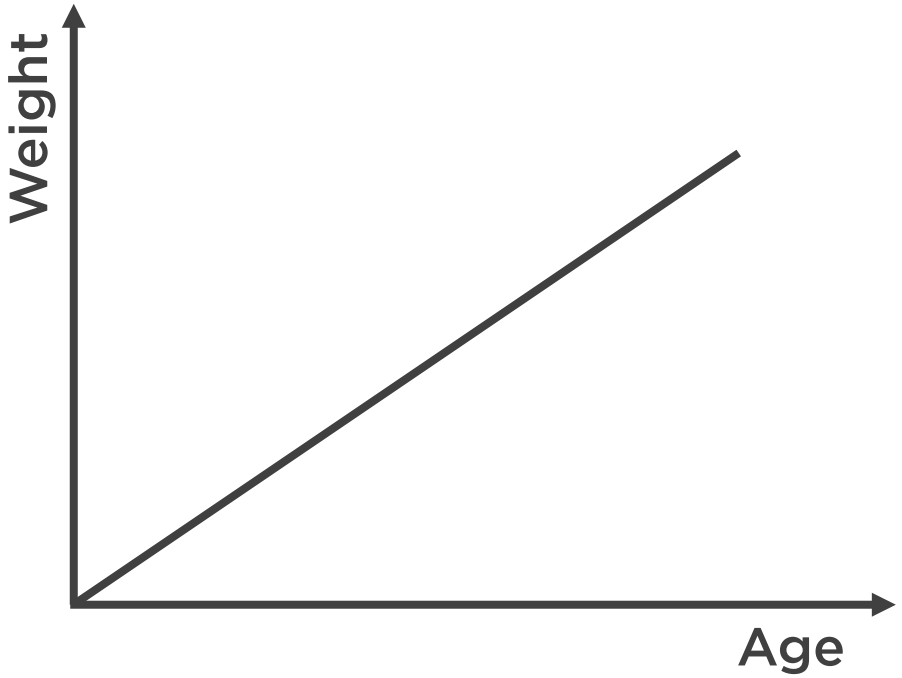
Building Linear Models



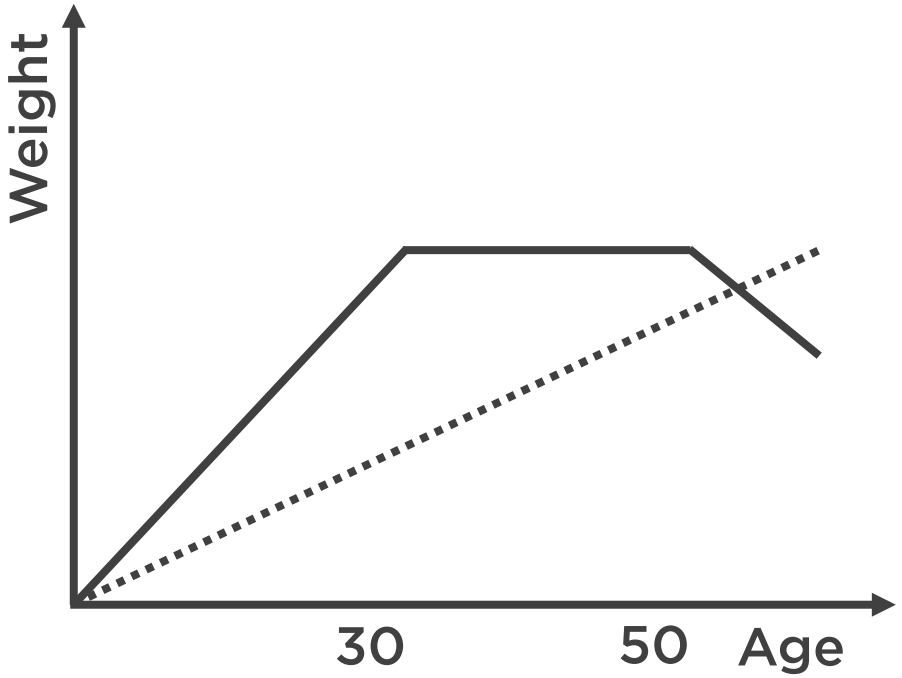
$$\text{Sale} = a_1(t) + a_2(t)^2 + a_3(t)^3 + C$$

Feature Engineering Numeric Variables

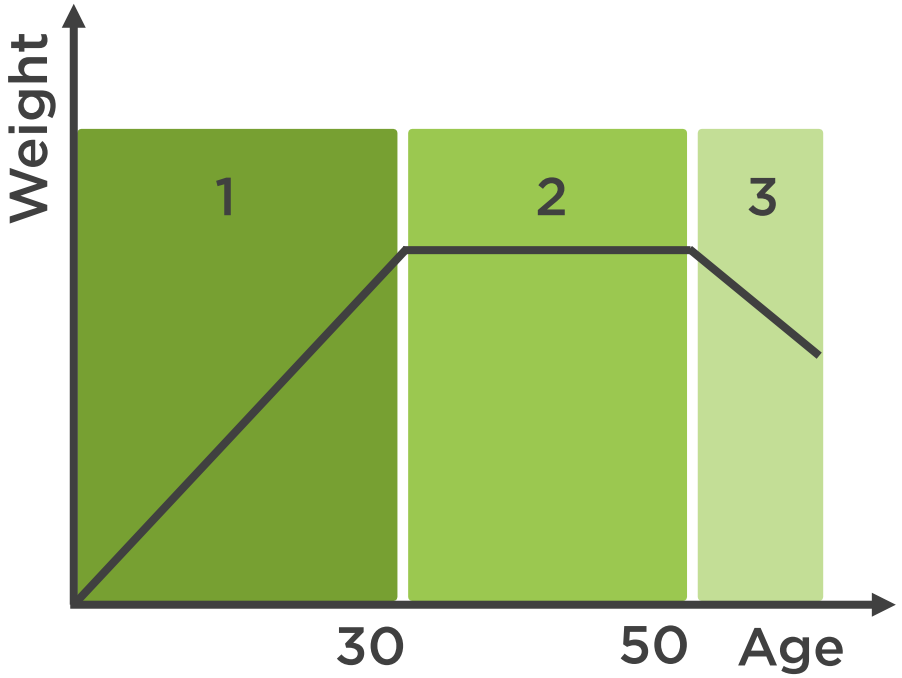
Why Binning?



Why Binning?



Why Binning?



Feature Engineering Categorical Variables

Predict Price of Apple

Small

Medium

Large

Predict Price of Apple

Size	Price
Small	XX
Medium	X
Small	X
Medium	XXX
Large	X
Large	X
Medium	X

Predict Price of Apple

Label Encoding

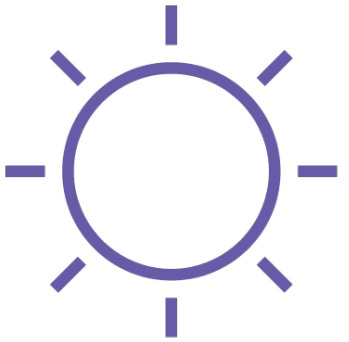
Size	Size Label	Price
Small	1	XX
Medium	2	X
Small	1	X
Medium	2	XXX
Large	3	X
Large	3	X
Medium	2	X

Predict Ice Cream Sales

Thunderstorm

Rainy

Snowy



Sunny

Predict Ice Cream Sales

Weather	Sales
Sunny	XX
Snowy	X
Rainy	X
Sunny	XXX
Rainy	X
Thunder	X
Snowy	X

Predict Ice Cream Sales

Weather	Weather_num	Sales
Sunny	1	XX
Snowy	2	X
Rainy	3	X
Sunny	1	XXX
Rainy	3	X
Thunder	4	X
Snowy	2	X

Predict Ice Cream Sales

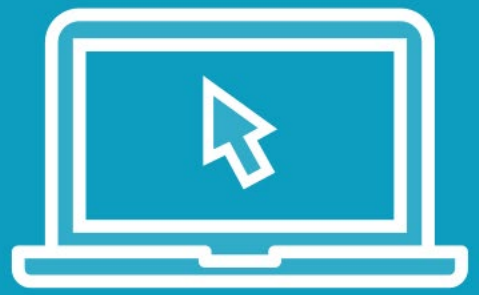
One-hot Encoding

Weather	Is_Snowy	Is_Sunny	Is_Rainy	Is_Thunder	Sales
Sunny	0	1	0	0	XX
Snowy	1	0	0	0	X
Rainy	0	0	1	0	X
Sunny	0	1	0	0	XXX
Rainy	0	0	1	0	X
Thunder	0	0	0	1	X
Snowy	1	0	0	0	X

Predict Ice Cream Sales

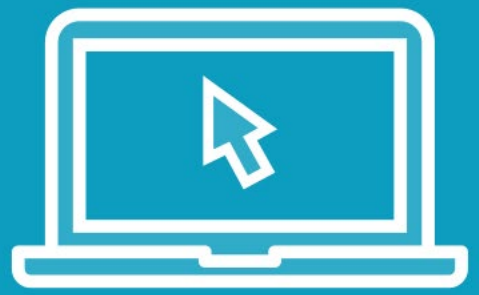
Weather	Is_Sunny	Is_Rainy	Is_Thunder	Sales
Sunny	1	0	0	XX
Snowy	0	0	0	X
Rainy	0	1	0	X
Sunny	1	0	0	XXX
Rainy	0	1	0	X
Thunder	0	0	1	X
Snowy	0	0	0	X

Demo



One-hot encoding

Demo



Learning with counts

Learning with Counts

Country	Income
United States	0
Mexico	1
United States	1
Mexico	0
United States	1
Germany	0
Mexico	0

Mexico	United States	Germany	Income
0	1	0	0
1	0	0	1
0	1	0	1
1	0	0	0
0	1	0	1
0	0	1	0
1	0	0	0

Learning with Counts

Country	Income
United States	0
Mexico	1
United States	1
Mexico	0
United States	1
Germany	0
Mexico	0

Class0_count	Class1_count
1	2
2	1
1	2
2	1
1	2
1	0
2	1

2/3 = 66 %

Summary

Feature Engineering is creating new features

More features increase model complexity

Feature Engineering helps to build simple models

Success of projects come down to features used