What is Statistical Learning?

Microsoft

Intuition behind Statistical Learning

N	Weight	Height	…	Age
1	166	153		59
2	177	167		49
3	159	176		50
4	169	150		75
5	148	161		53
6	179	180		47
7	153	169		51
8	156	157		89
9	162	158		44

?

Blood Pressure
?
?
?
?
?
?
?
?
?

http://faculty.marshall.usc.edu/gareth-james/ISL/

Traditional Approach

N	Weight	Height	…	Age	$Y = (0.3 * Weight) + (0.25 * Height) + (0.5 * Age)$	Blood Pressure
1	166	153		59		
2	177	167		49		
3	159	176		50		
4	169	150		75		
5	148	161		53		
6	179	180		47		
7	153	169		51		
8	156	157		89		
9	162	158		44		

Traditional Approach

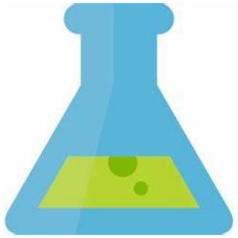
N	Weight	Height	…	Age	$Y = (0.3 * Weight) + (0.25 * Height) + (0.5 * Age)$	Blood Pressure
1	166	153		59	----------→	117.55
2	177	167		49	----------→	119.35
3	159	176		50	----------→	116.7
4	169	150		75	----------→	125.7
5	148	161		53	----------→	111.15
6	179	180		47	----------→	122.2
7	153	169		51	----------→	113.65
8	156	157		89	----------→	130.55
9	162	158		44	----------→	110.1

Statistical Learning Approach

N	Weight	Height	…	Age		Blood Pressure
1	166	153		59		117.55
2	177	167		49		119.35
3	159	176		50		116.7
4	169	150		75		125.7
5	148	161		53		111.15
6	179	180		47		122.2
7	153	169		51		113.65
8	156	157		89		130.55
9	162	158		44		110.1

Statistical Learning Approach

N	Weight	Height	…	Age	Blood Pressure
1	166	153		59	117.55
2	177	167		49	119.35
3	159	176		50	116.7
4	169	150		75	125.7
5	148	161		53	111.15
6	179	180		47	122.2
7	153	169		51	113.65
8	156	157		89	130.55
9	162	158		44	110.1



$$\boldsymbol{Y = f(x)}$$

$$\boldsymbol{Y = (0.3 * Weight) + (0.25 * Height) + (0.5 * Age)}$$

Intuition behind Statistical Learning

https://www.akkencloud.com/5-ways-handle-negative-comments-on-social-media/

Microsoft

Intuition behind Statistical Learning

https://www.akkencloud.com/5-ways-handle-negative-comments-on-social-media/

Microsoft

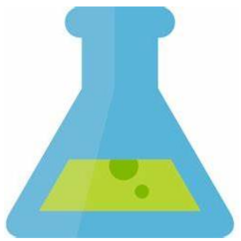
Why do we estimate f(x)?

Microsoft

Why do we use Statistical Learning?

Prediction

N	Weight	Height	…	Age	Blood Pressure
1	166	153		59	117.55
2	177	167		49	119.35
3	159	176		50	116.7
4	169	150		75	125.7
5	148	161		53	111.15
6	179	180		47	122.2
7	153	169		51	113.65
8	156	157		89	130.55
9	162	158		44	110.1



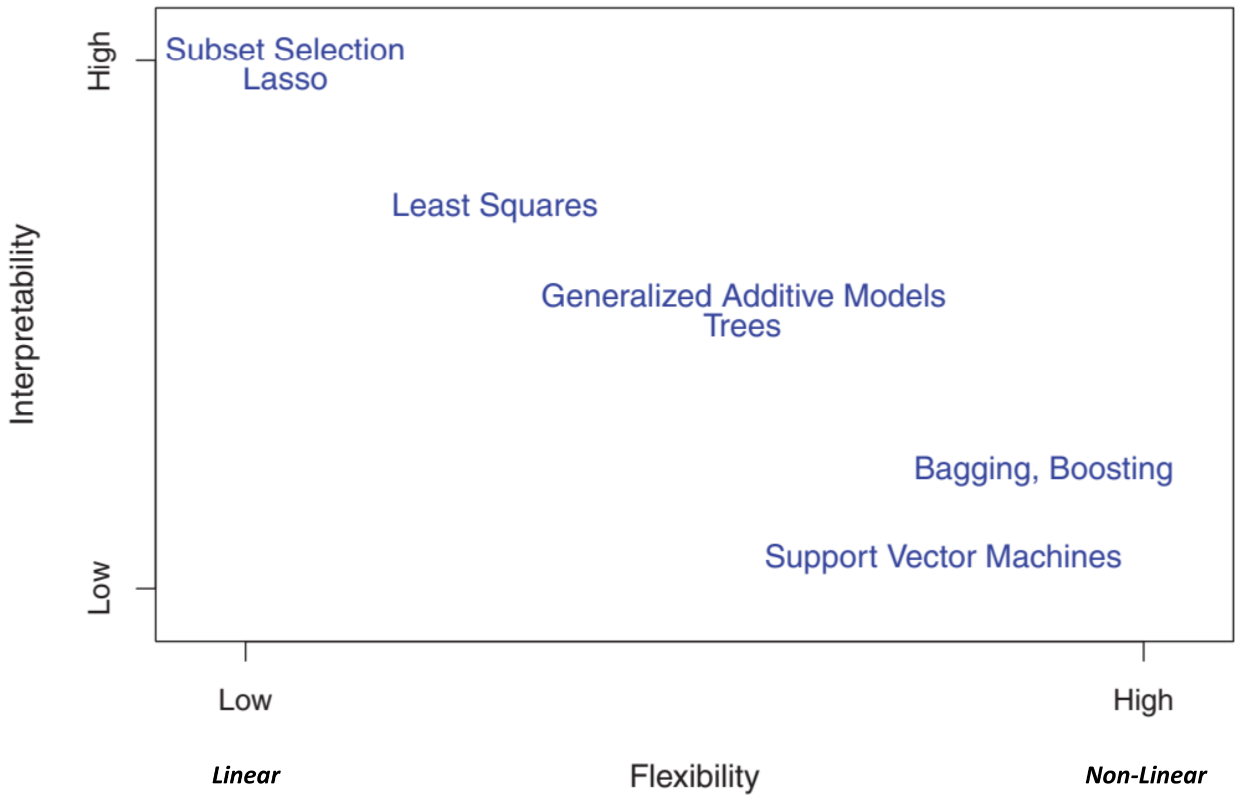
$$\boldsymbol{Y = f(x)}$$

$$\boldsymbol{Y = (0.3 * Weight) + (0.25 * Height) + (0.5 * Age)}$$

Why do we use Statistical Learning?

- Which ***Predictors*** are associated with the ***response***?
  - Not all variables used in models are substantially associated to the response.
- What is the ***relationship*** between the response and each ***predictor***?
  - Some predictors can have positive effects in the response. Other predictors may have opposite effects.
- Can the relationship between ***Y*** and each predictor be adequately summarized using a ***linear equation***, or is the relationship more complicated?
  - Not always the nature of the relationship is linear.

Trade-Off Between Prediction Accuracy and Model Interpretability



- The more linear the model, the more explainable.
- The more non-linear (flexible) the more

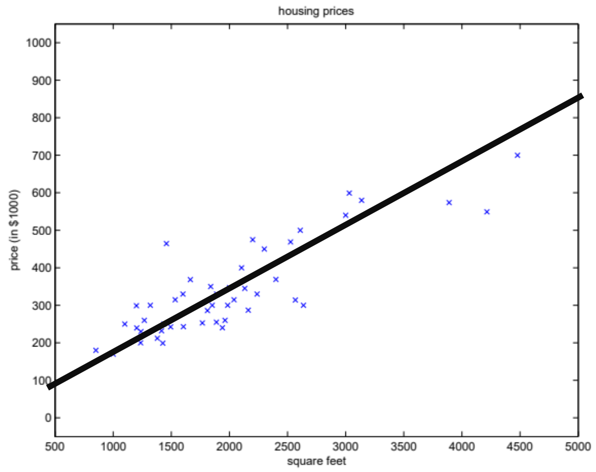
http://faculty.marshall.usc.edu/gareth-james/ISL/

Microsoft

Types of Machine Learning Problems

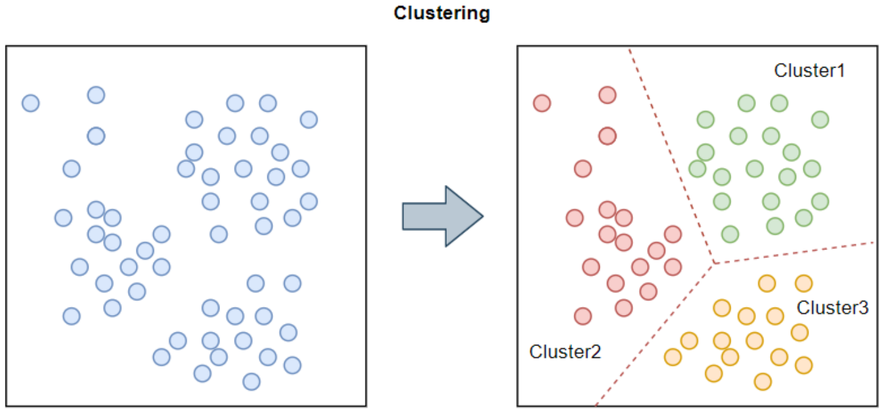
Supervised Learning

Living area ($feet^2$)	Price (1000$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	



We know the input (***features/predictors***) and the response or ***target***.

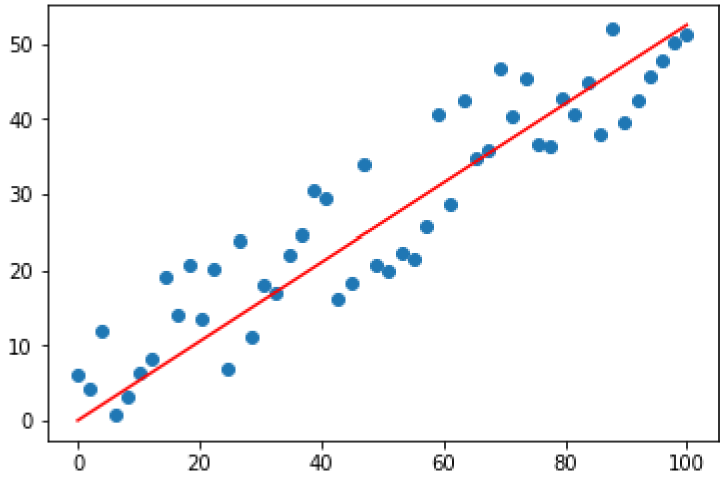
Unsupervised Learning



We do not have the answer (***Target***). However, we want to find structure in data (***clustering***).

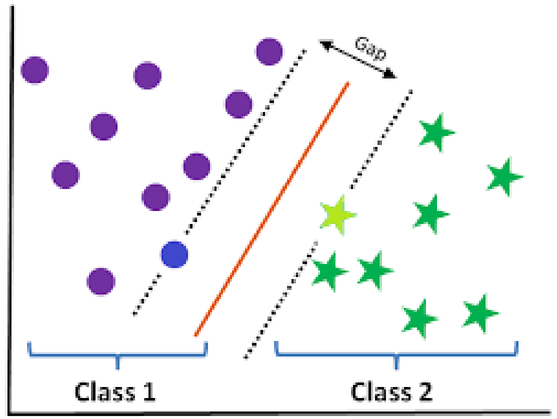
Supervised Learning

Regression



We are trying to ***predict*** a value (***Sales, Blood pressure***)

Classification

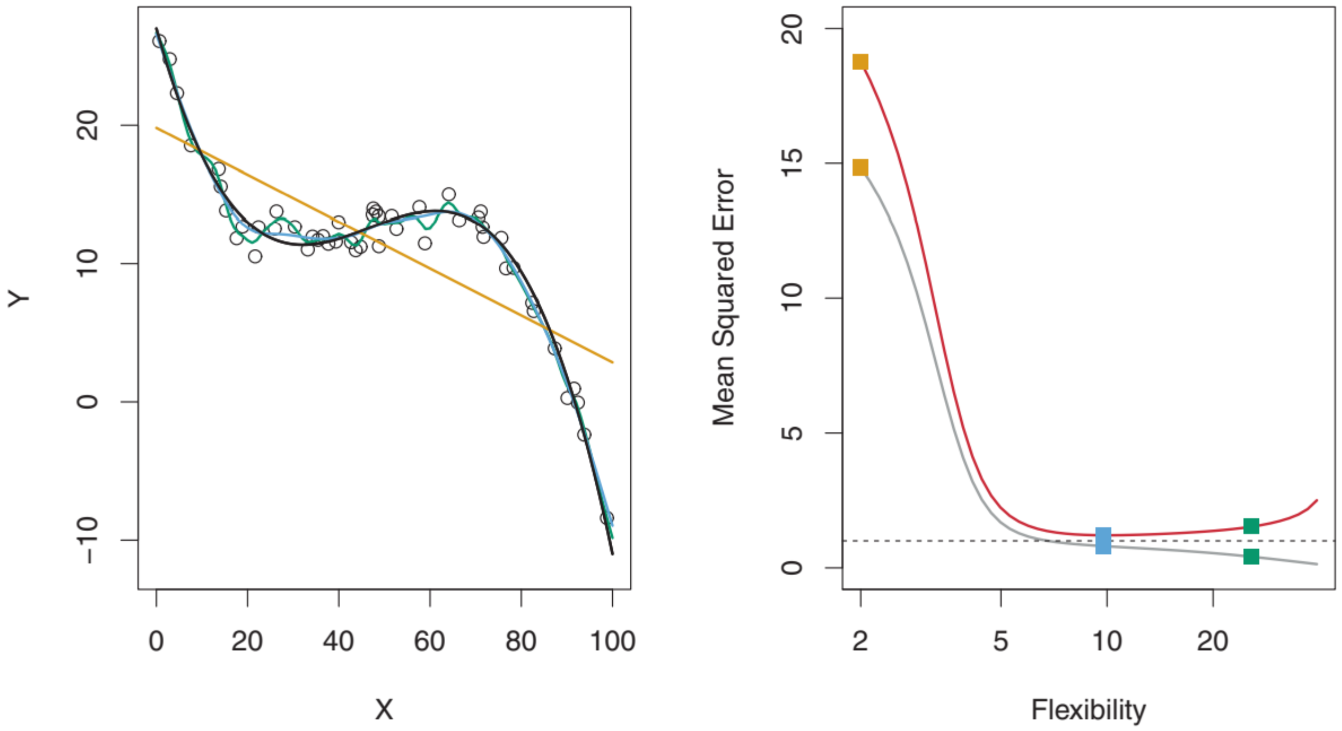


We are trying to ***predict*** whether a specific point belongs to a ***category***

http://faculty.marshall.usc.edu/gareth-james/ISL/

Different types of Regression

Error vs. Flexibility



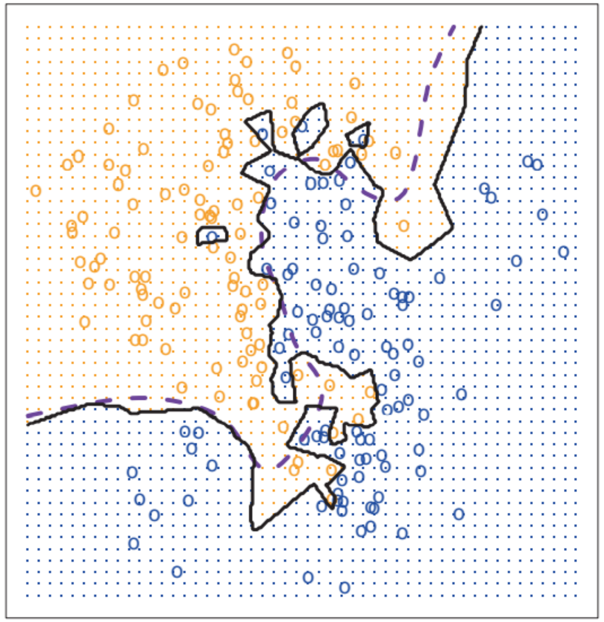
* In the left chart we can find that we can adjust different types of curves to the same data. The nature of those curves can be either linear or non-linear

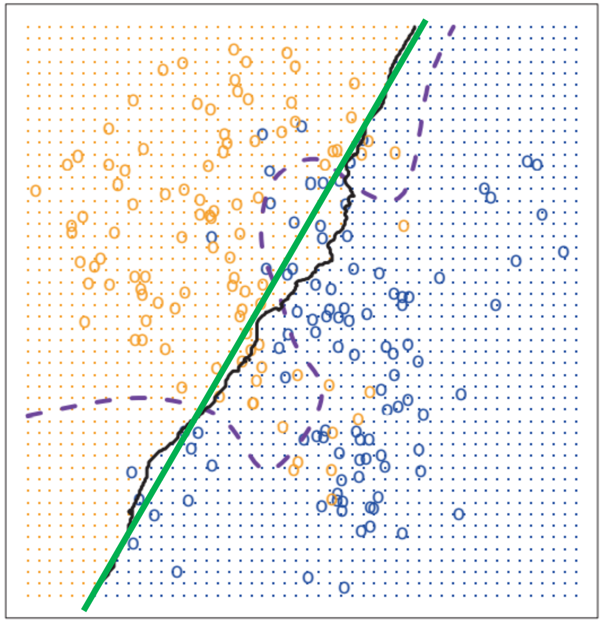
http://faculty.marshall.usc.edu/gareth-james/ISL/

Microsoft

Different types of Classification

Classification





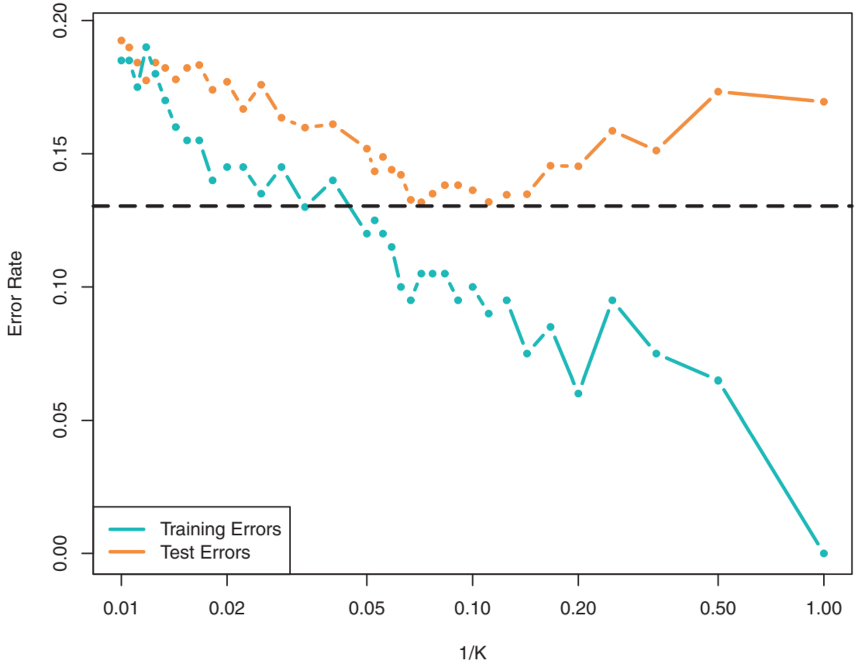
*In the left chart we can find that we can adjust different types of curves to the same data. The nature of those curves can be either linear or non-linear

http://faculty.marshall.usc.edu/gareth-james/ISL/

Microsoft

Different types of Classification

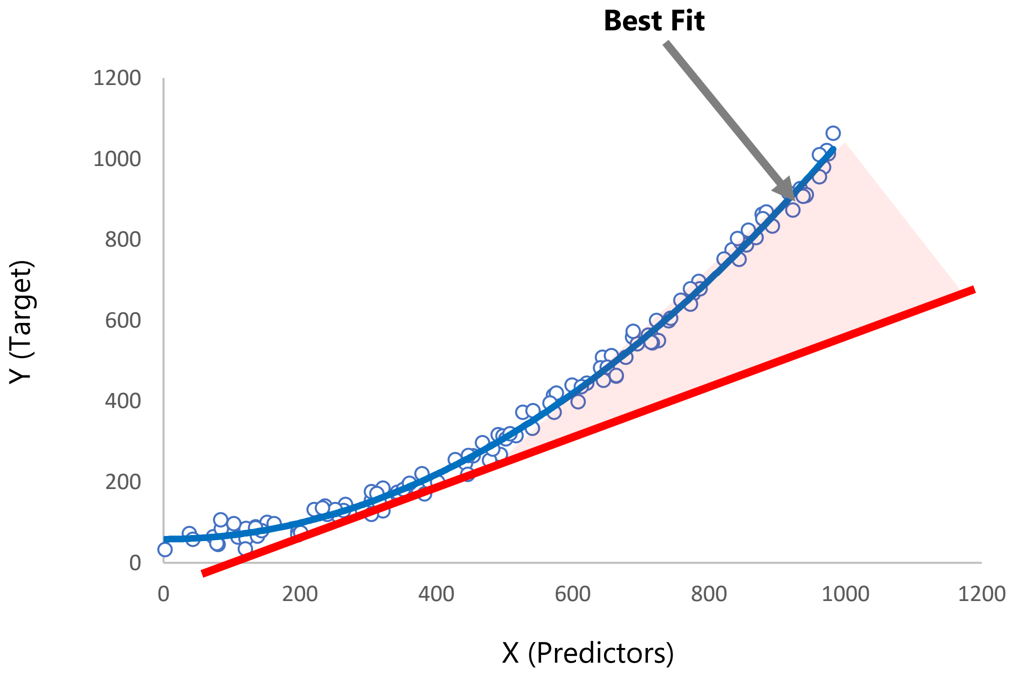
Classification

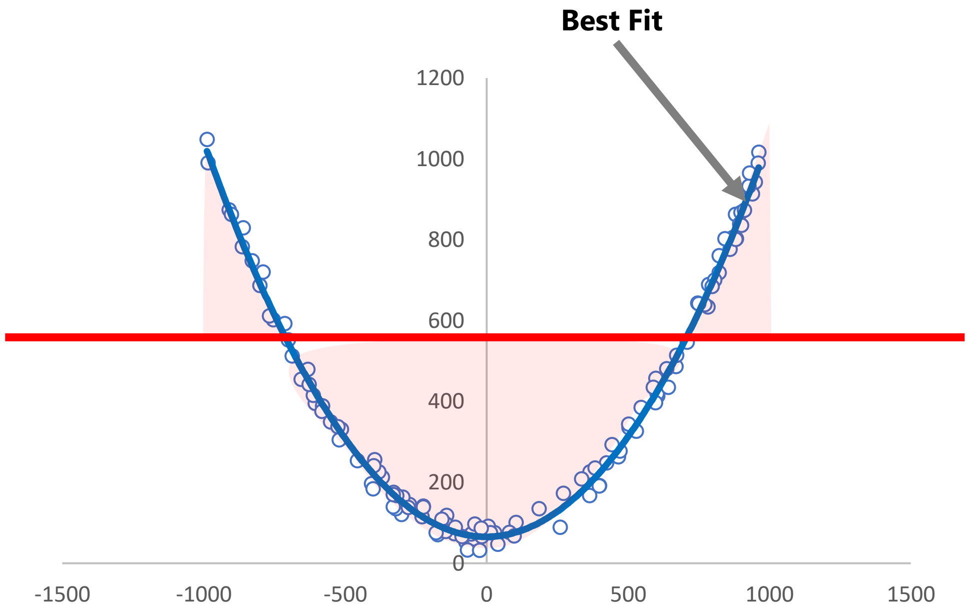


The more specific (**more flexible the model**), the higher the expected error for the testing data (**new data**).

http://faculty.marshall.usc.edu/gareth-james/ISL/

Bias



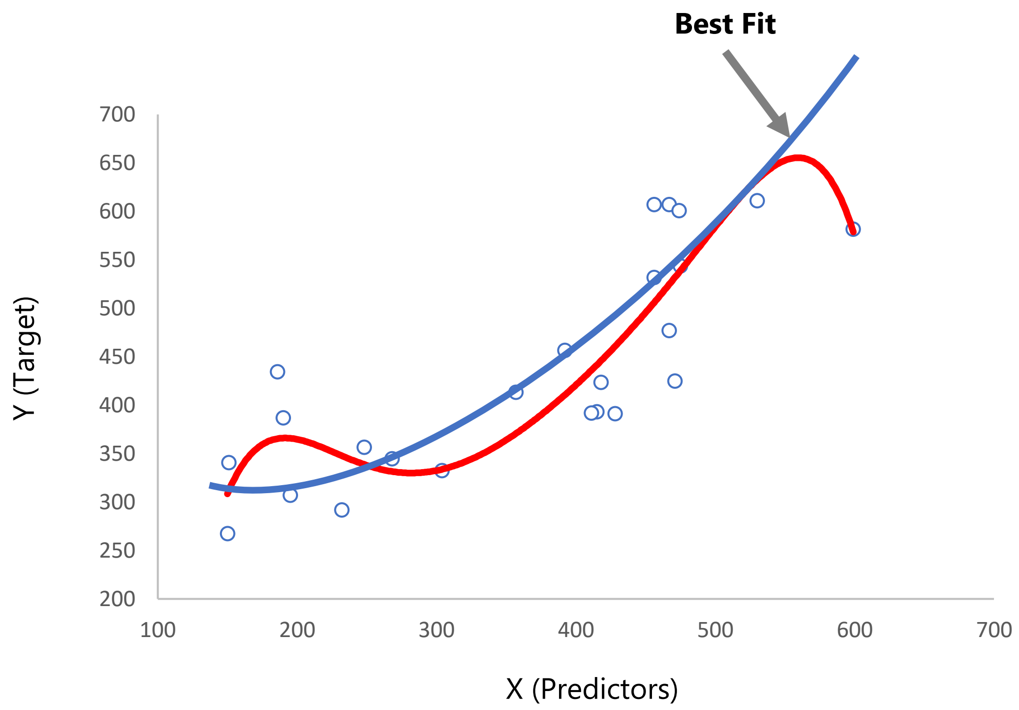


Bias can be understood as the behavior of always **underrepresenting** the data. ***Bias*** refers to the error of trying to explain an extremely complicated problem (***data***) with a much simpler model.

http://faculty.marshall.usc.edu/gareth-james/ISL/

Microsoft

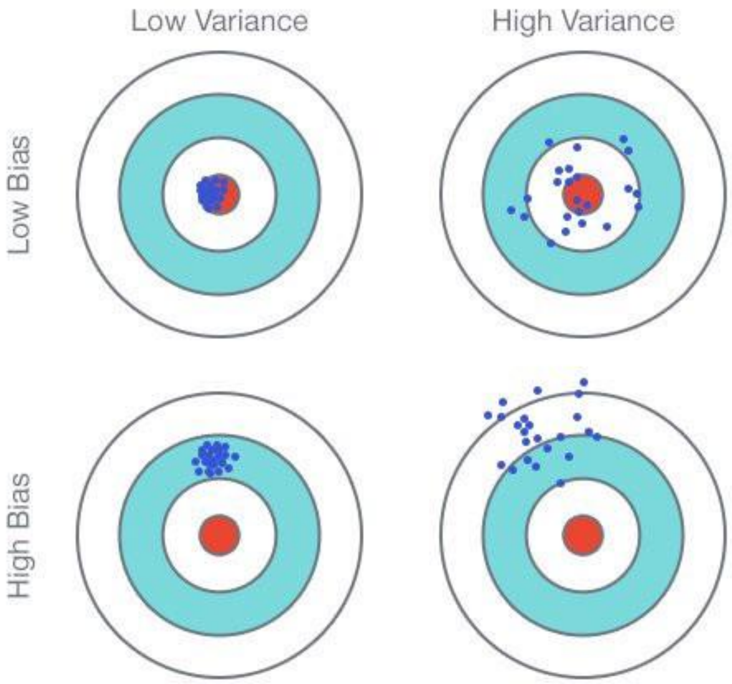
Variance



Variance can be understood as overrepresenting the training data. This means that the model is too specific to the data that the model has seen. However, it would not perform very well on new data.

http://faculty.marshall.usc.edu/gareth-james/ISL/

Bias and Variance

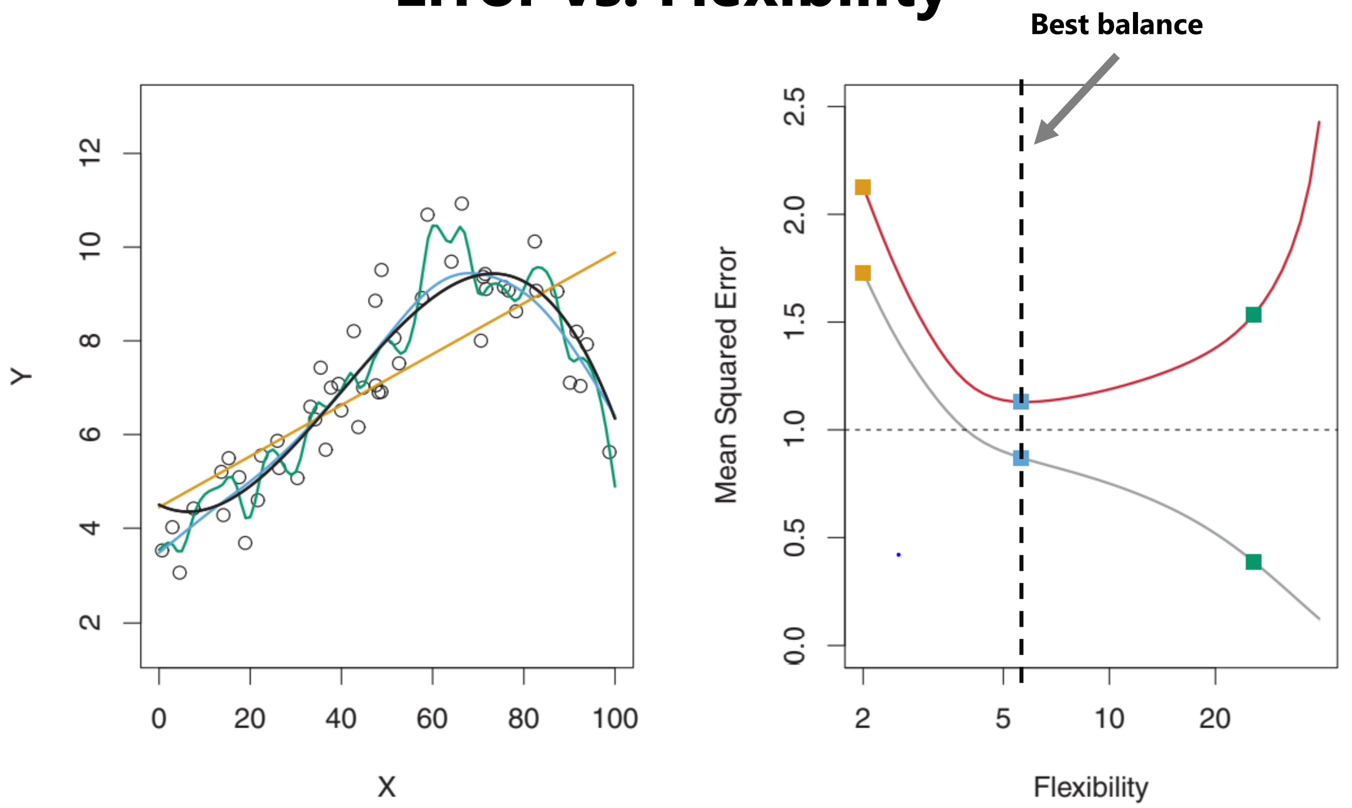


Visual representation of **Bias** and **Variance**.

http://faculty.marshall.usc.edu/gareth-james/ISL/

Bias-Variance Trade-off

Error vs. Flexibility



The total error of the model is the sum of both the **Bias** and **Variance**. Thus we always want find the right balance between them.

http://faculty.marshall.usc.edu/gareth-james/ISL/

Microsoft