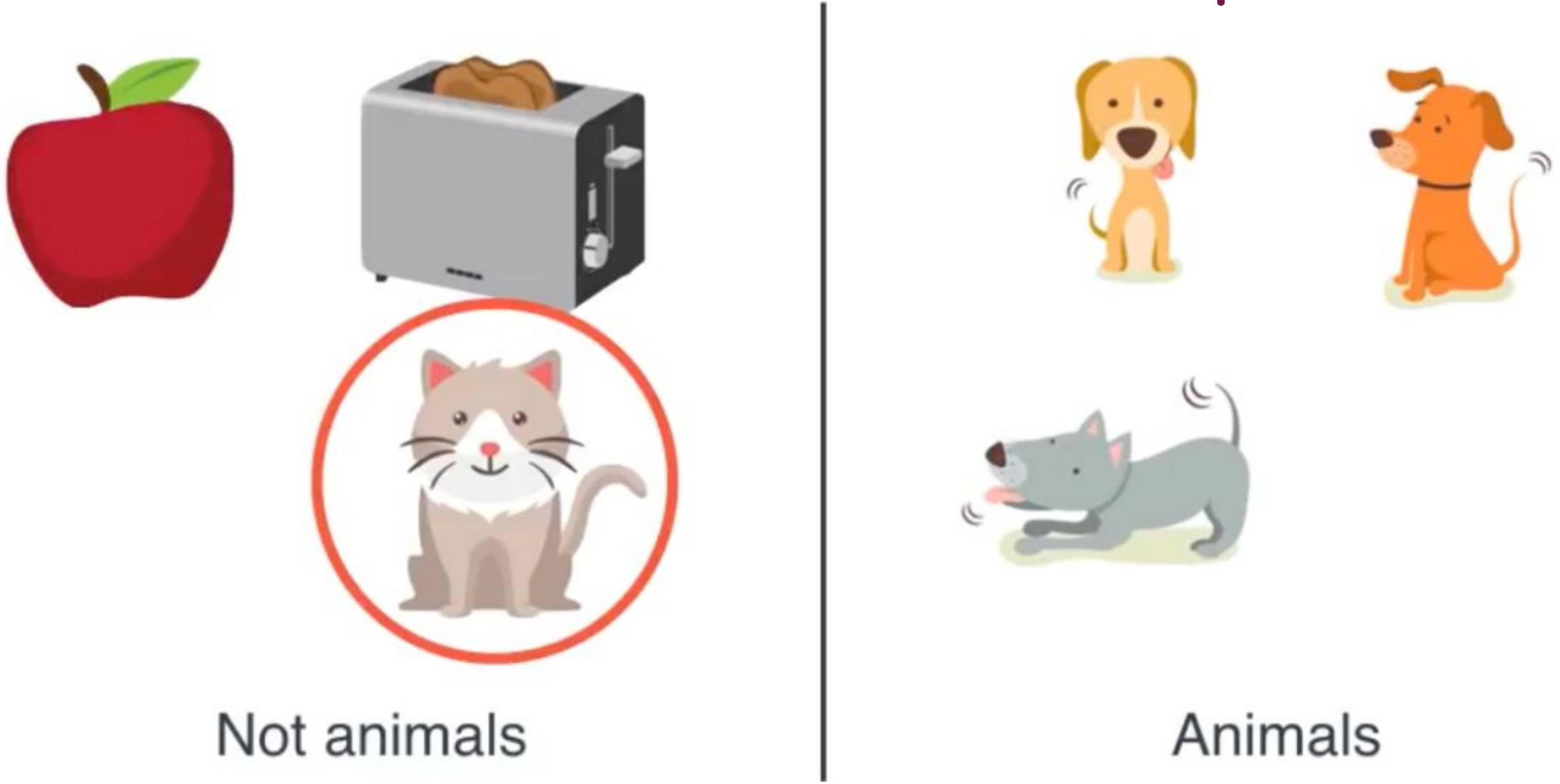
Bias Variance

Prepared By: Dr.Mydhili K Nair, Professor, ISE Dept, RIT
For: Machine Learning Elective Class
Target Audience: Sem 6 Students
Term: Feb to June 2019

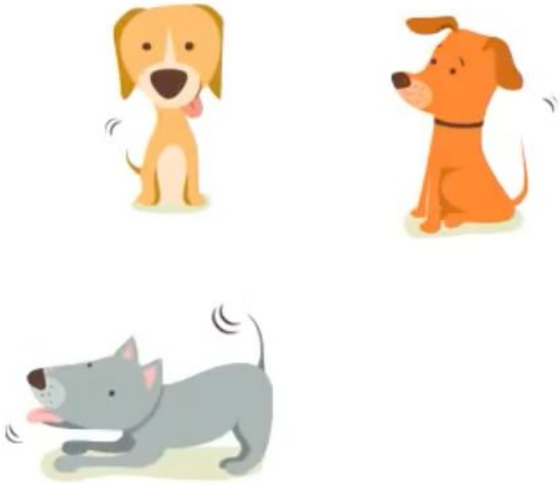
Bias

Our Classification Model is too simple



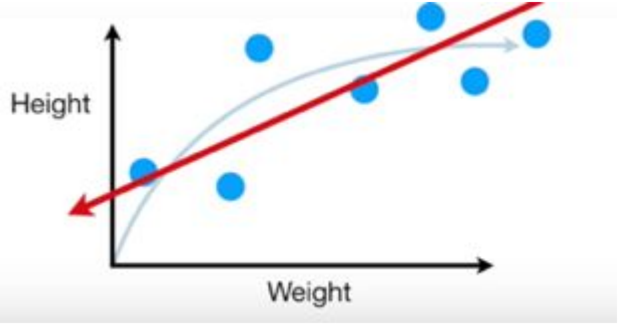
Our Classification Model is too specific

Anything but dogs that are wagging their tail

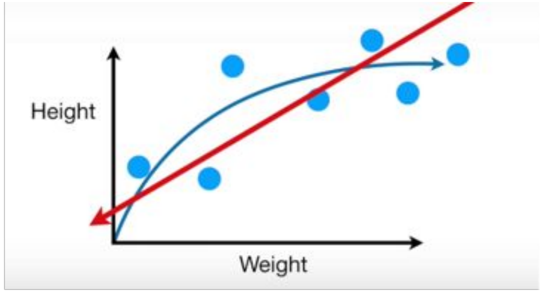


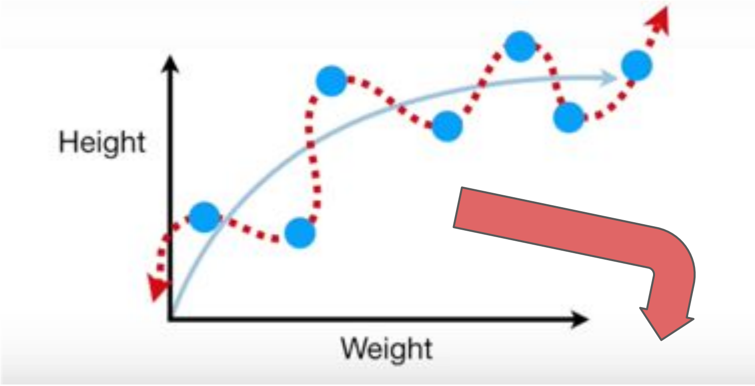
Too specific

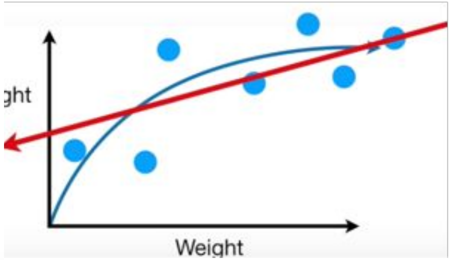
Dogs that are wagging their tail



Linear Model: How much ever you vary the "prediction regression line" it will not fit the curve and there will always be a **bias** between actual value and predicted value.



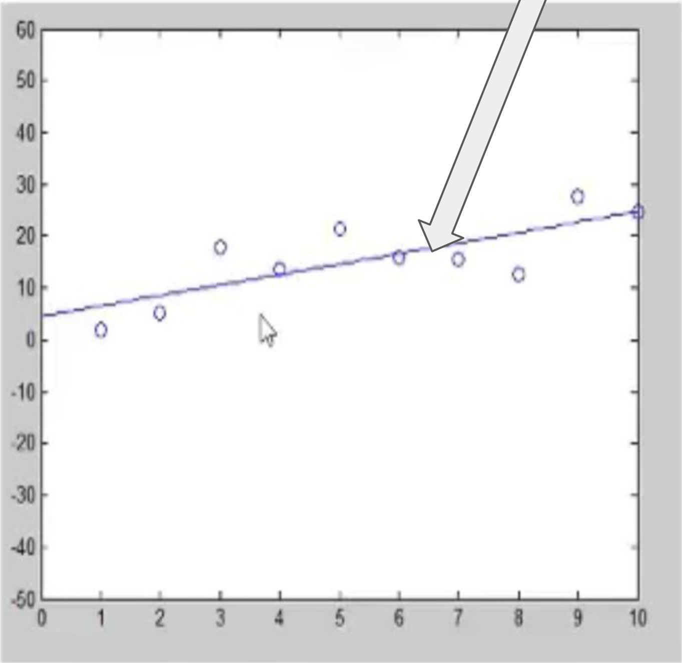




Bias error is completely eliminated. This non-linear model fits the data points perfectly. Zero Bias.

Model 1: $y = b_0 + b_1 x$

Linear Model

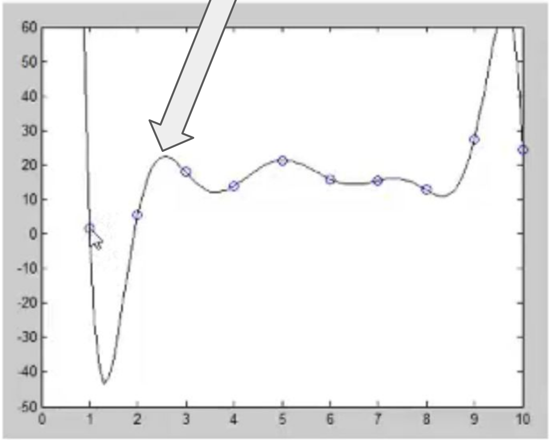


x	y
1	1.7
2	5.3
3	18.0
4	13.8
5	21.4
6	15.9
7	15.5
8	12.7
9	27.5
10	24.6

Model 2: $y = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \cdots + b_9 x^9$

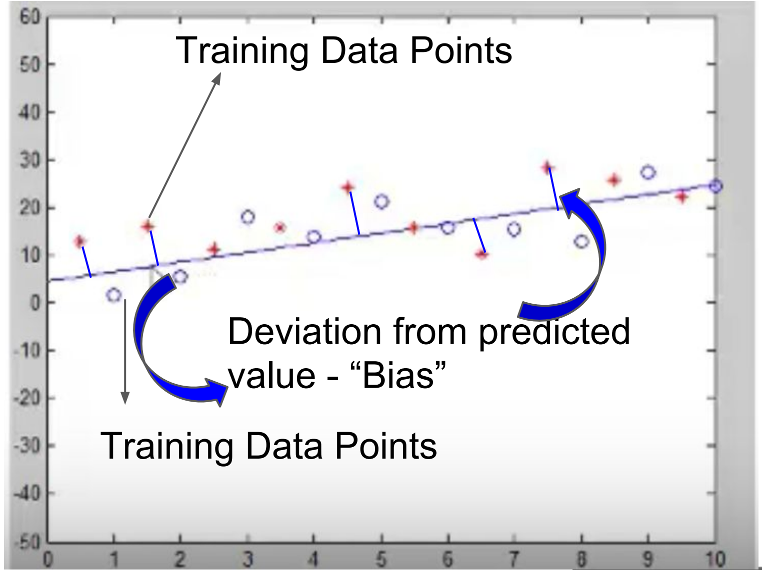
x	x^2		x^9	y
1	1	…	1	1.7
2	4		262144	5.3
3	9		4E+08	18.0
	16		7E+10	13.8
	25		4E+12	21.4
	36		1E+14	15.9
	49		2E+15	15.5
	64		2E+16	12.7
	81		2E+17	27.5
	100		1E+18	24.6

9th Order Polynomial Model



$$\text{Model 1: } y = b_0 + b_1 x$$

Our Classification Model is too simple - HIGH BIAS

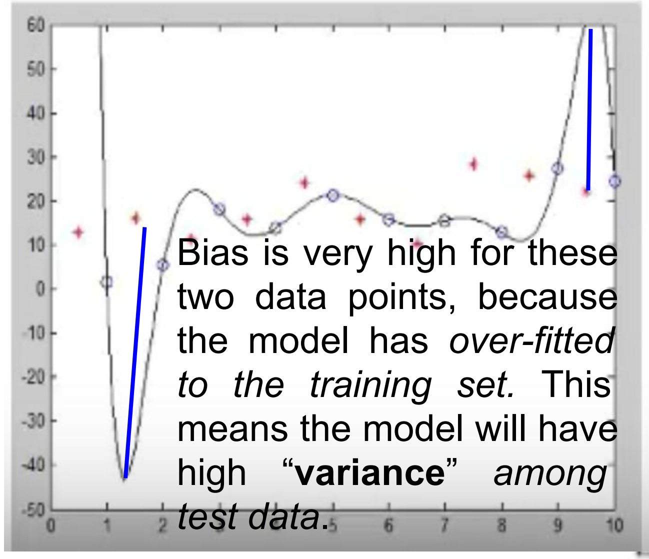


x	y
1	1.7
2	5.3
3	18.0
4	13.8
5	21.4
6	15.9
7	15.5
8	12.7
9	27.5
10	24.6

$$\text{Model 2: } y = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \cdots + b_9 x^9$$

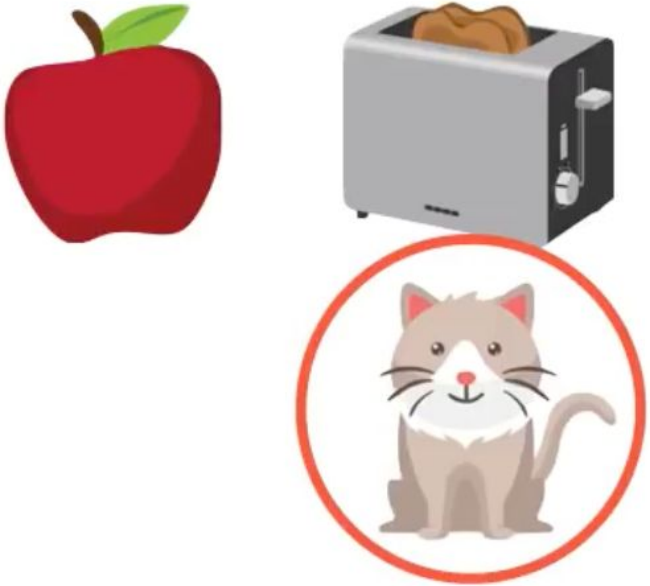
x	x^2		x^9	y
1	1	...	1	1.7
2	4		262144	5.3
3	9		4E+08	18.0
4	16		7E+10	13.8
5	25		4E+12	21.4
6	36		1E+14	15.9
7	49		2E+15	15.5
8	64		2E+16	12.7
9	81		2E+17	27.5
10	100		1E+18	24.6

Our Classification Model is too specific - **HIGH VARIANCE**



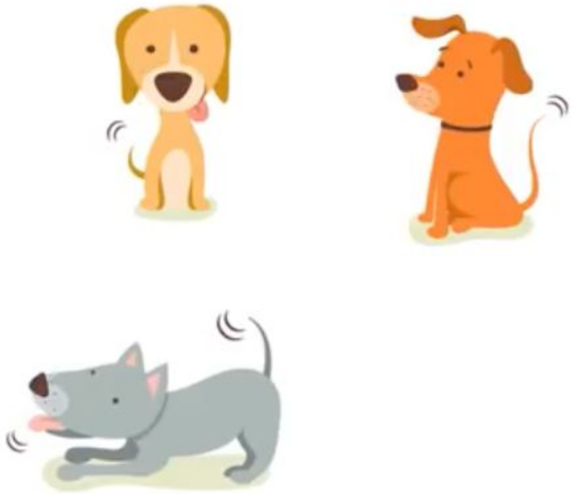
Bias is very high for these two data points, because the model has *over-fitted to the training set.* This means the model will have high "**variance**" *among test data.*

Error due to bias (underfitting)



Not animals

Too simple

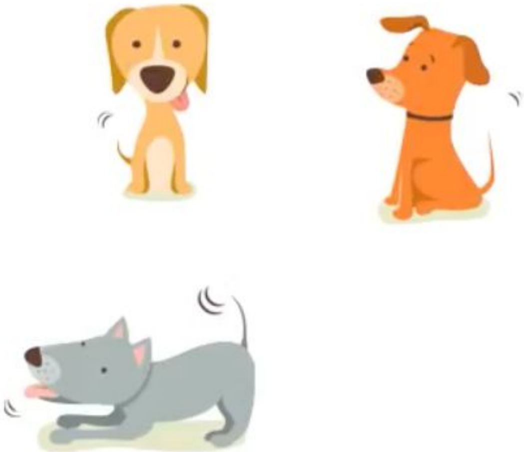


Animals

Error due to variance (overfitting)

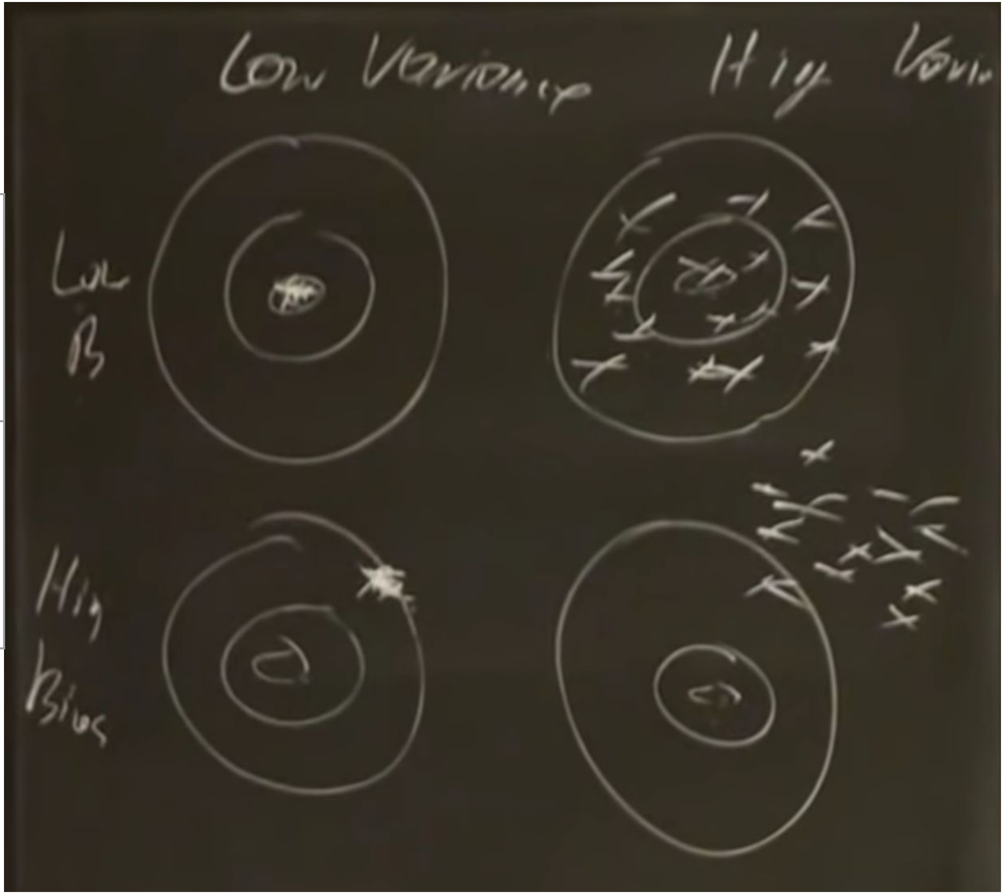
Anything but dogs that are
wagging their tail

Too specific



Dogs that are wagging their tail

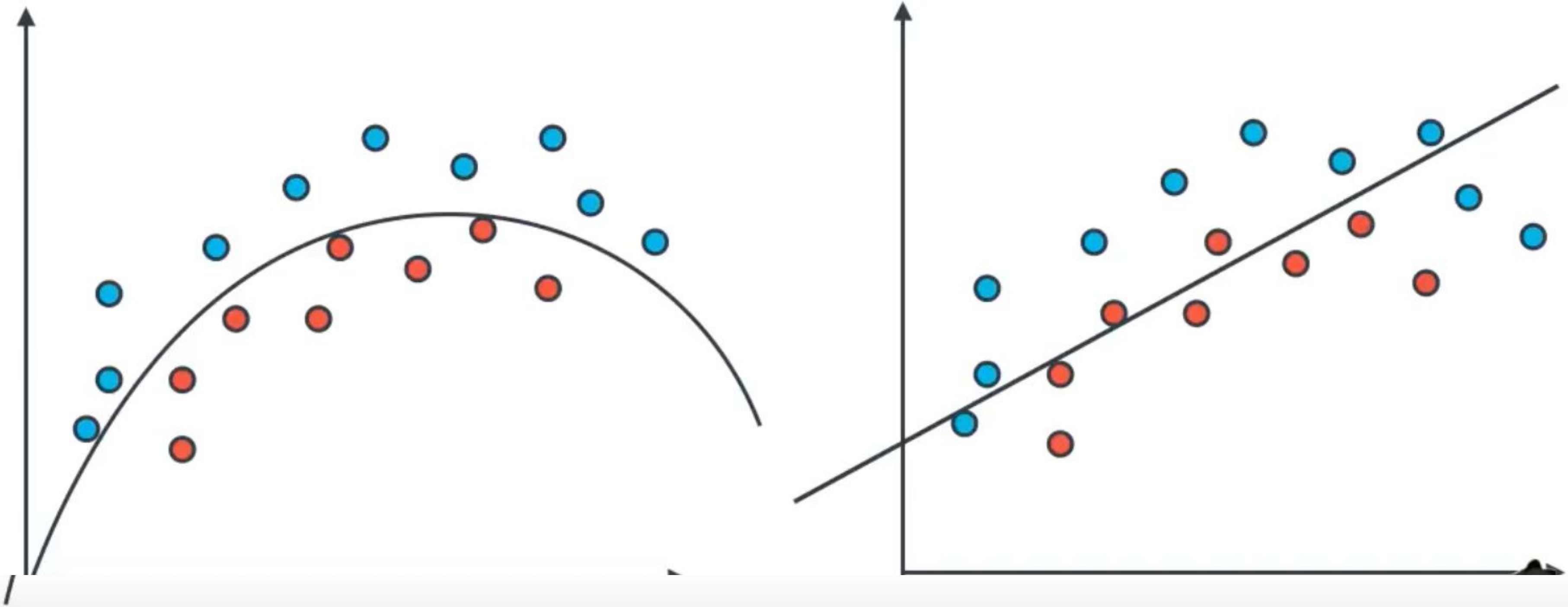
	Low Variance	High Variance
Low Bias		
High Bias		



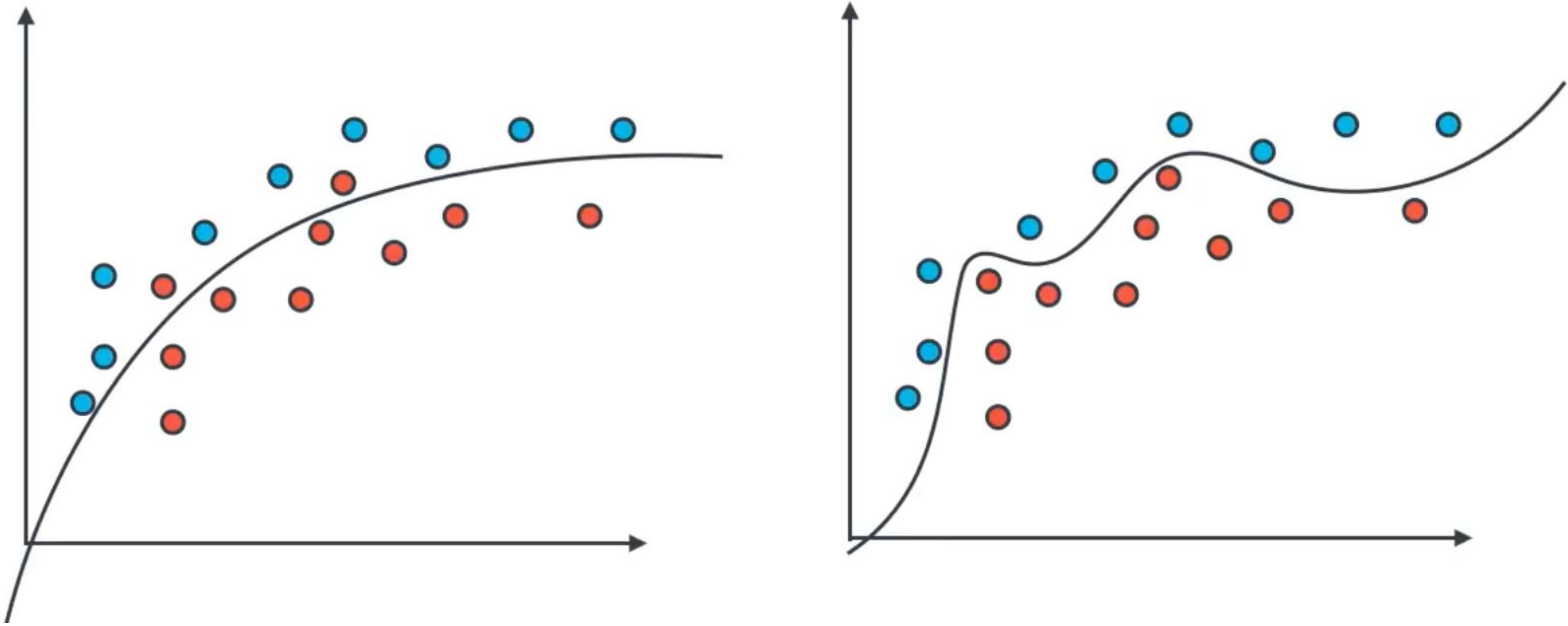
Every Model aims at *Low Bias & Low Variance*

Bias - Variance Tradeoff

Error due to bias (underfitting)



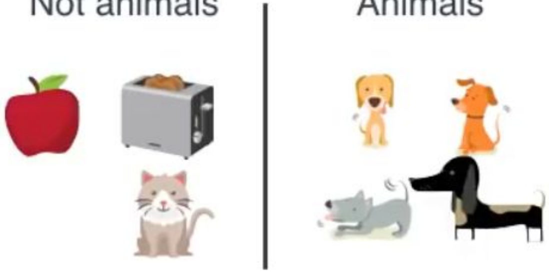
Error due to variance (overfitting)



Tradeoff

High bias (Underfitting)

Not animals | Animals



Bad on Training set

Bad on Testing set

Just Right

Not dogs | Dogs

Good on Training set

Good on Testing set

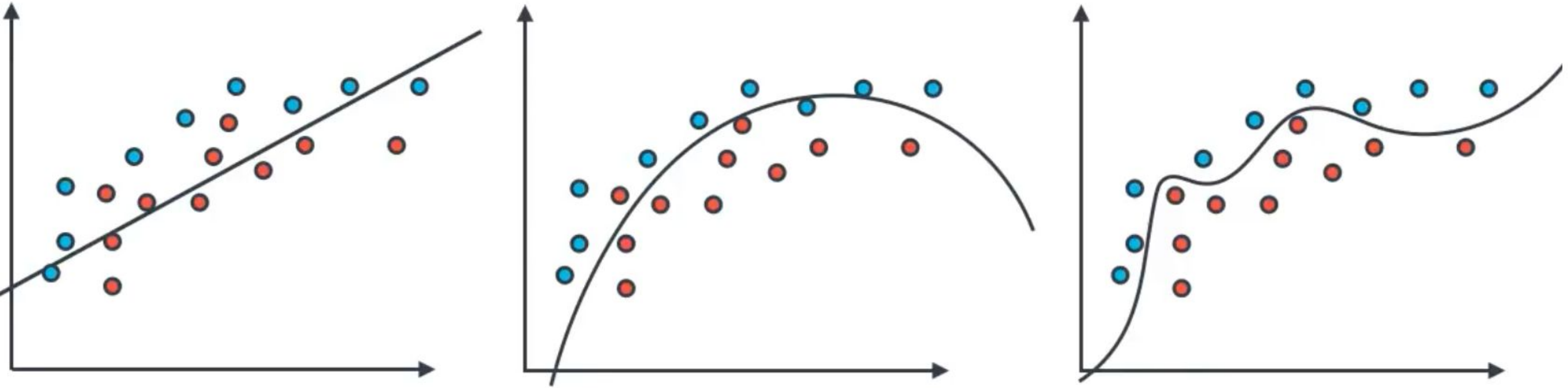
High variance (Overfitting)

Not dogs who wag their tail | Dogs who wag their tail

Great on Training set

Bad on Testing set

Model Complexity Graph



High Bias

degree = 1

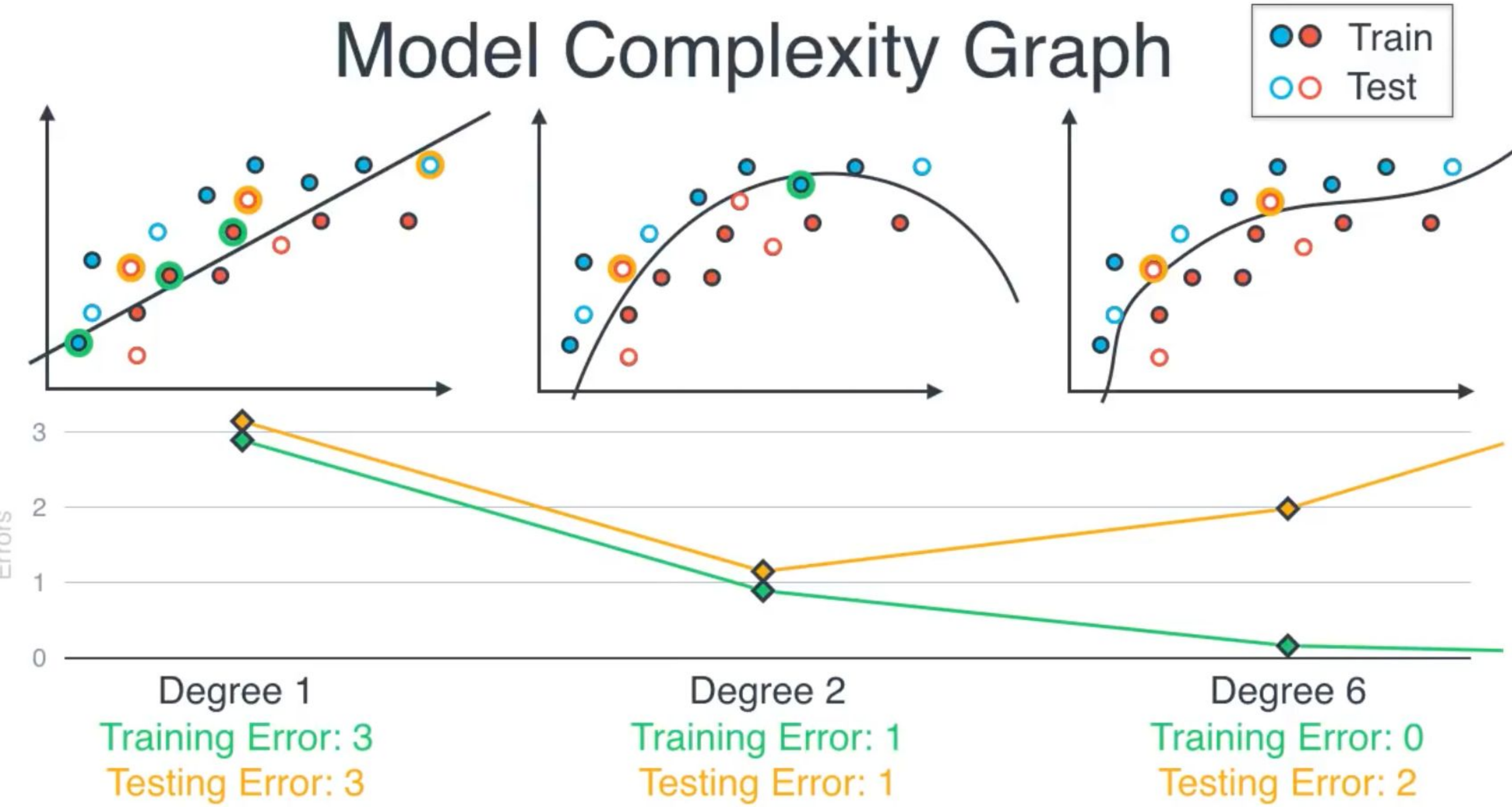
Just Right

degree = 2

High Variance

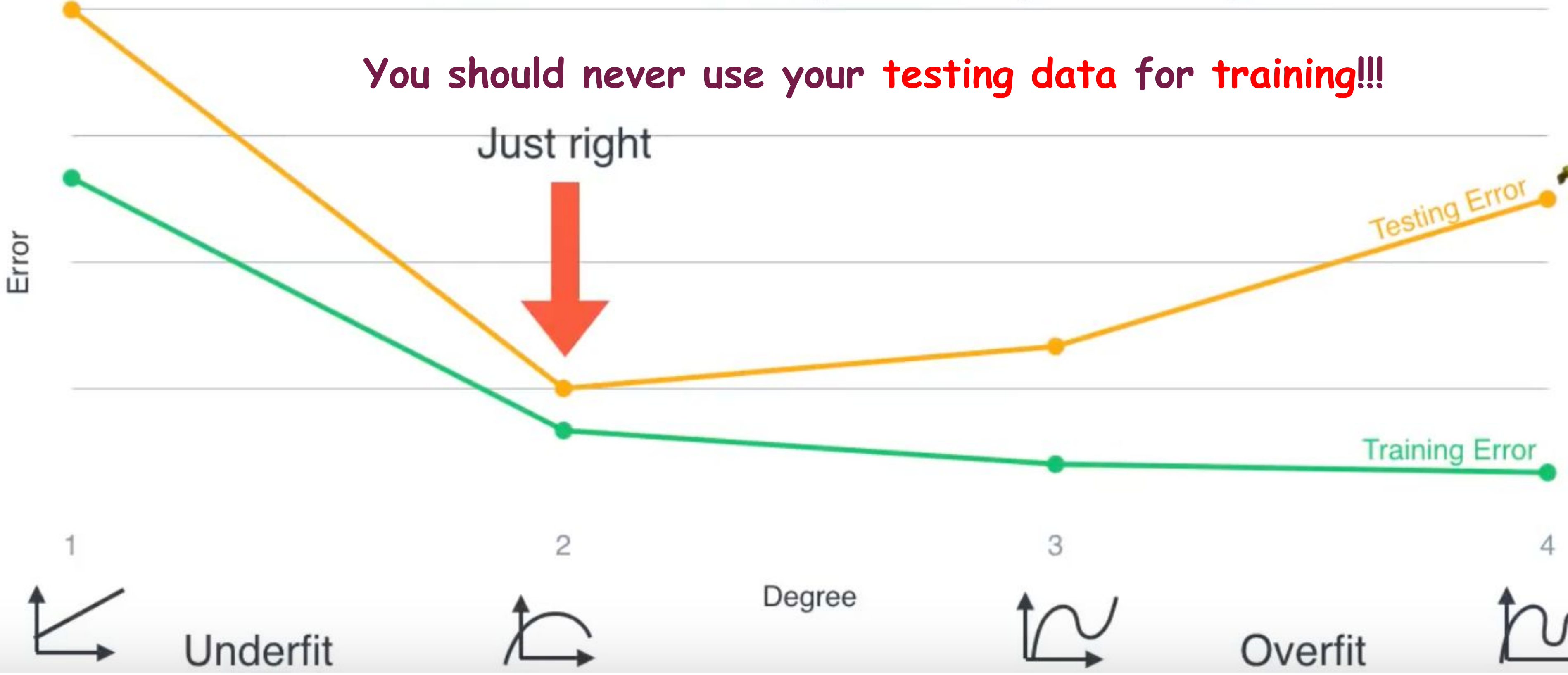
degree = 6

Model Complexity Graph



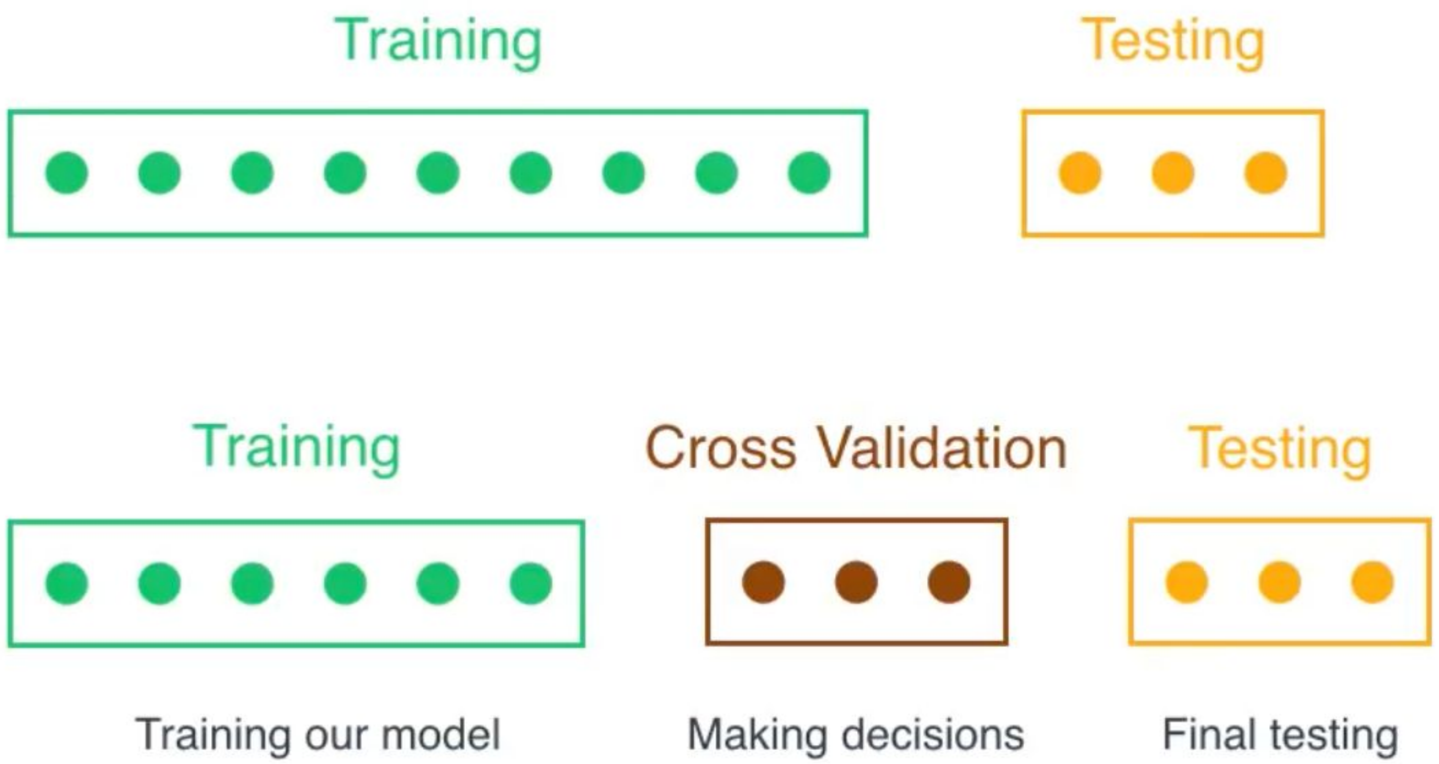
Model Complexity Graph

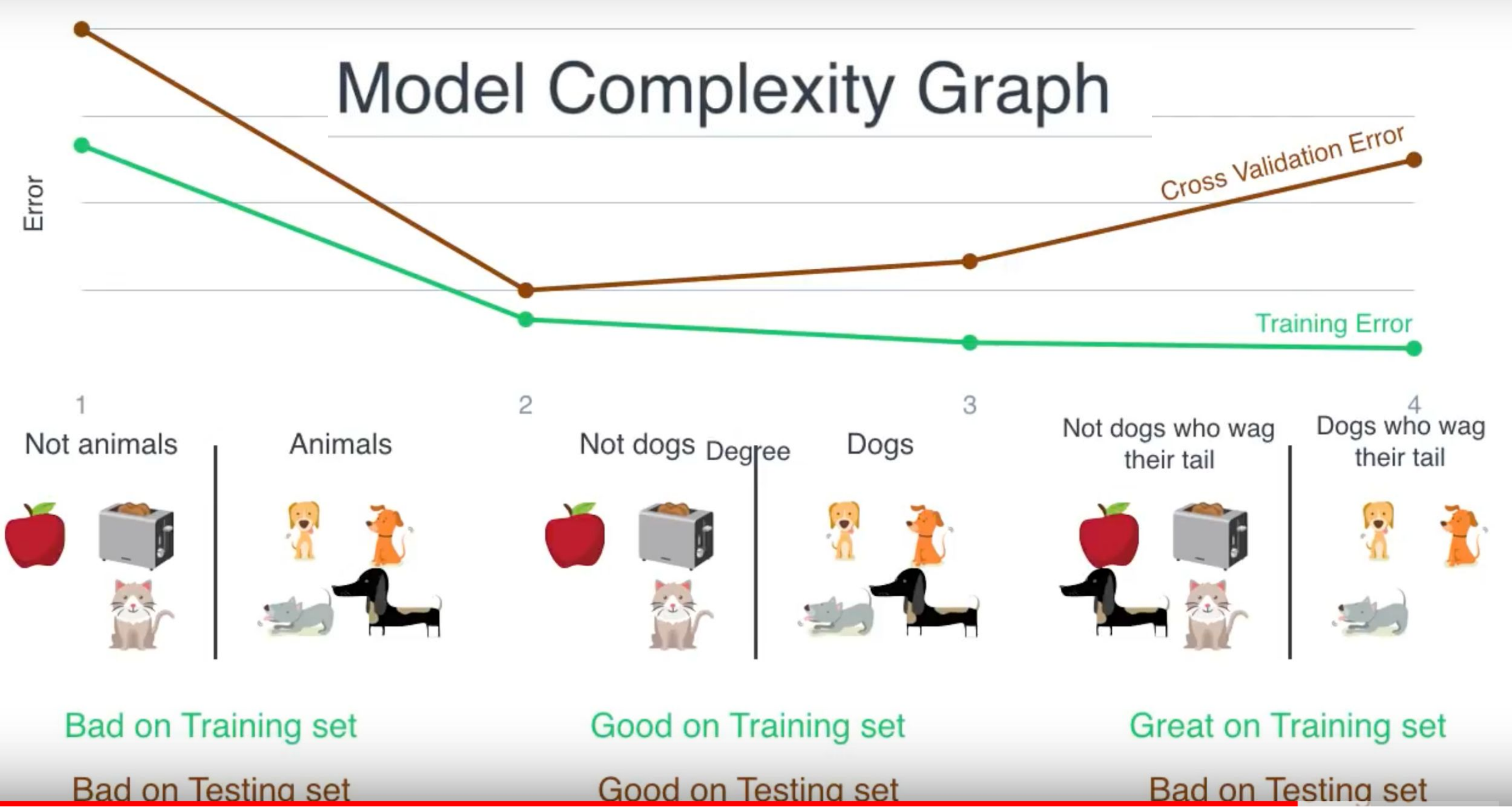
You should never use your testing data for training!!!



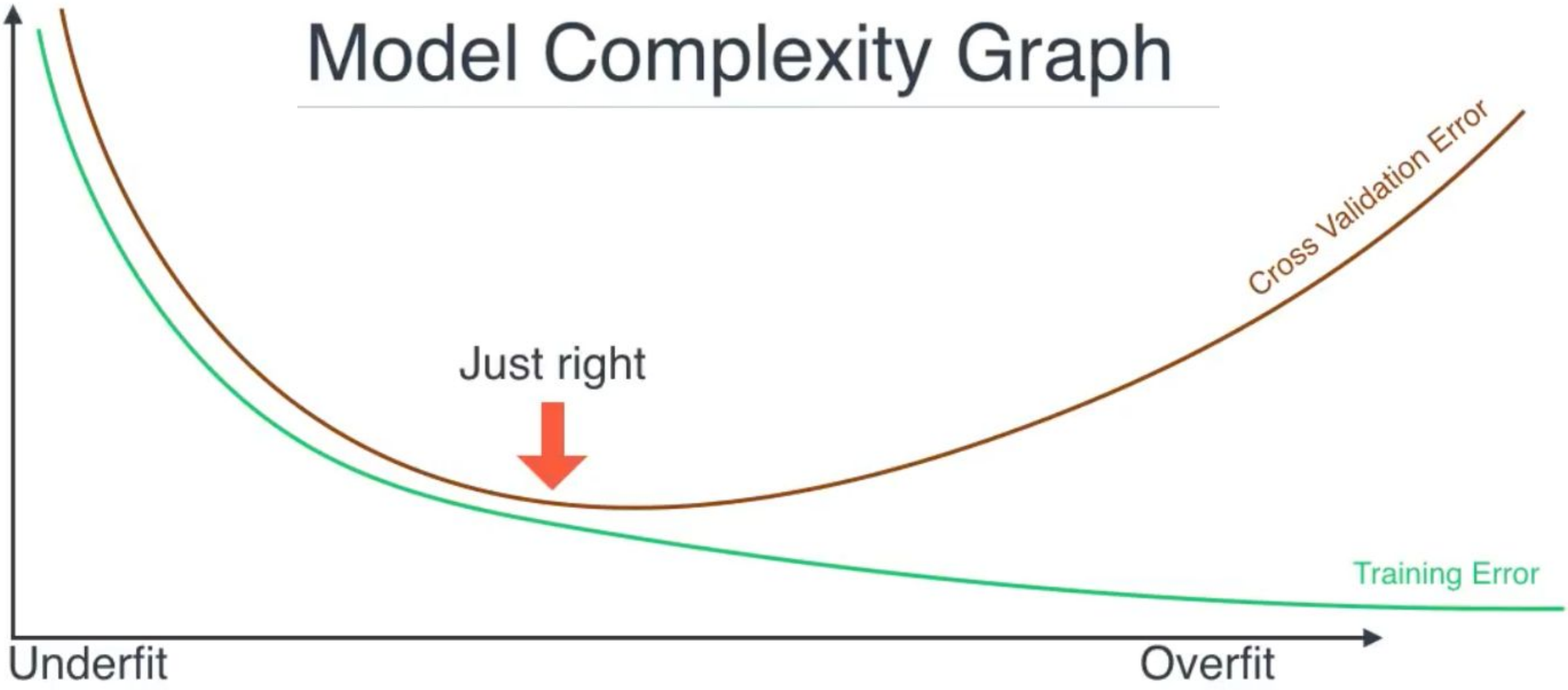
Training - Testing - Validation Datasets

Solution: Cross Validation

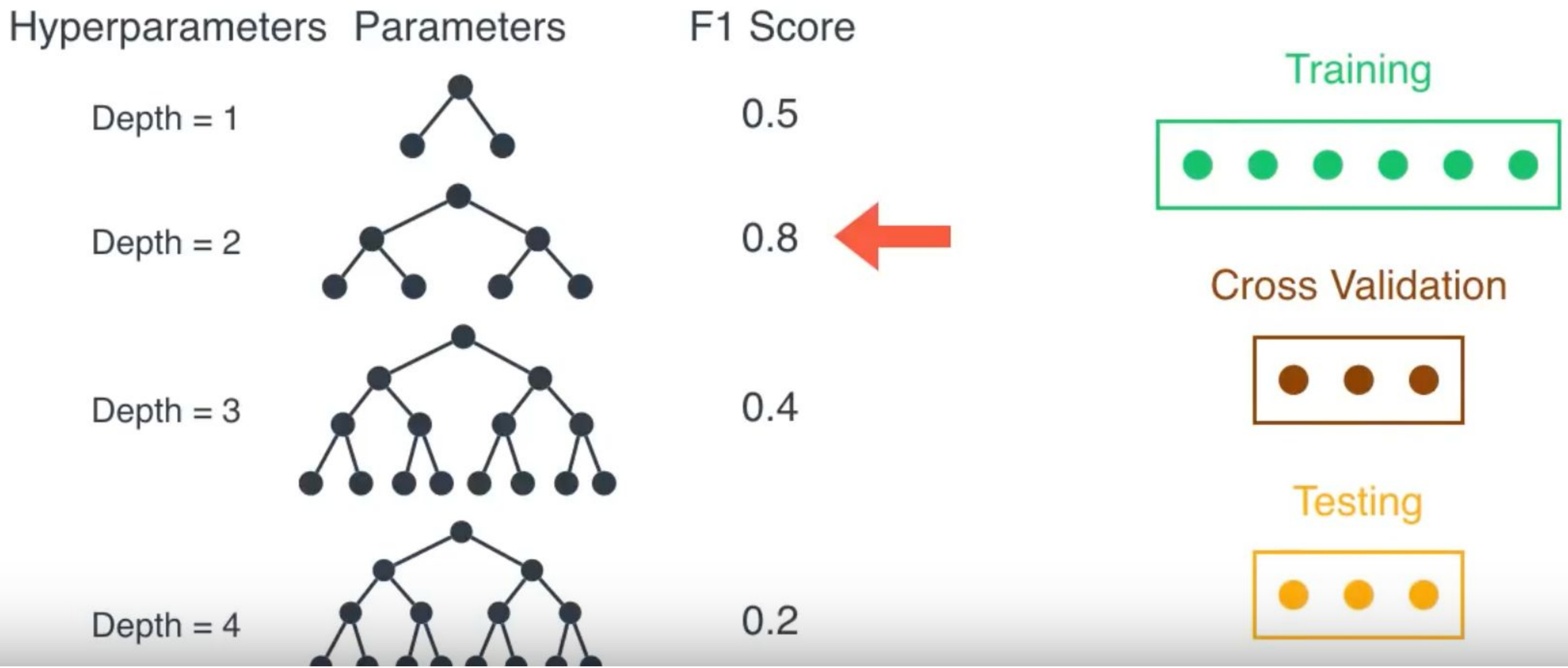



Model Complexity Graph
Error
Cross Validation Error
Training Error
1
2
3
4
Degree
Not animals
Animals
Not dogs
Dogs
Not dogs who wag their tail
Dogs who wag their tail
Bad on Training set
Bad on Testing set
Good on Training set
Good on Testing set
Great on Training set
Bad on Testing set

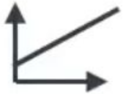
Model Complexity Graph

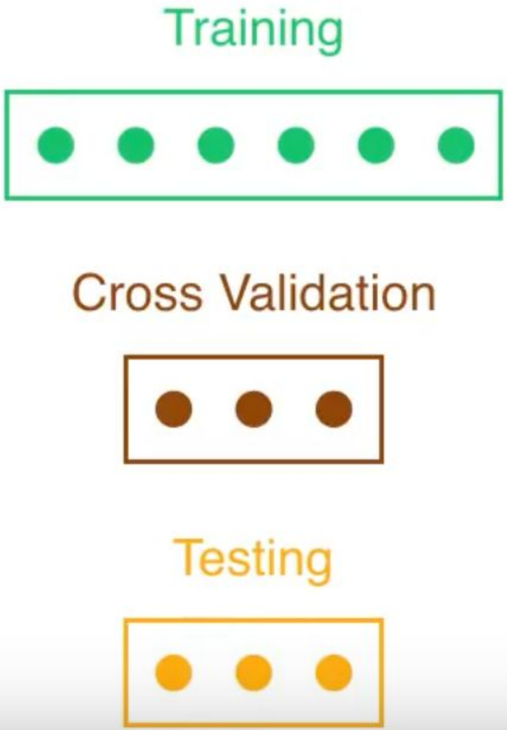


Training a Decision Tree



Training a Logistic Regression Model

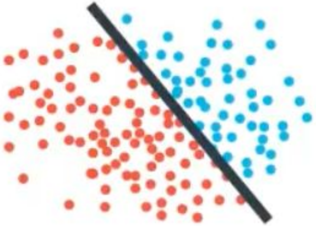
	Parameters	F1 Score
Degree = 1		0.5
Degree = 2		0.8
Degree = 3		0.4
Degree = 4		0.2



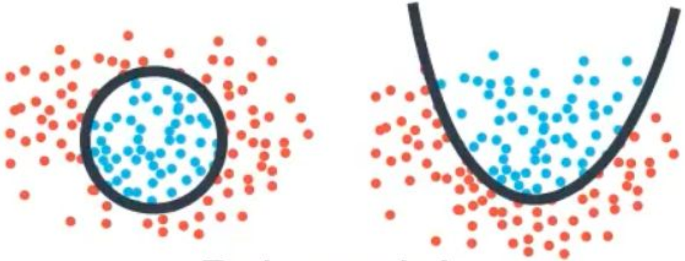
Training a Support Vector Machine

Hyperparameters

Kernel

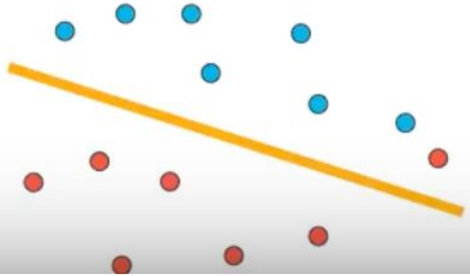


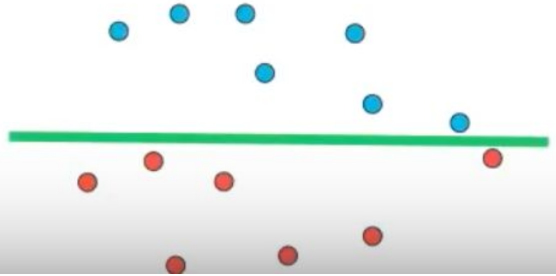
Linear



Polynomial

γ

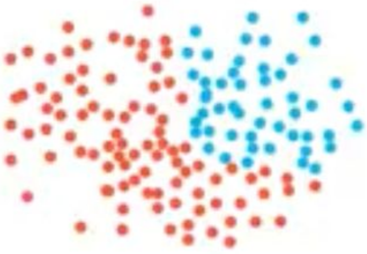




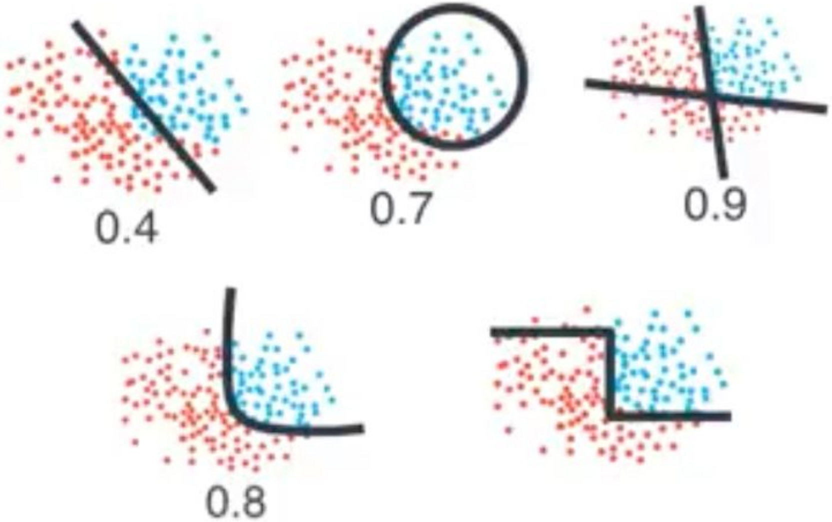
Parameters and Hyperparameters

Algorithm	Parameters	Hyperparameters
Random Forest	Features Thresholds	Number of trees Depth
Logistic Regression	Coefficients of the polynomial	Degree of the polynomial
Support Vector Machines	Coefficients	Kernel Gamma C

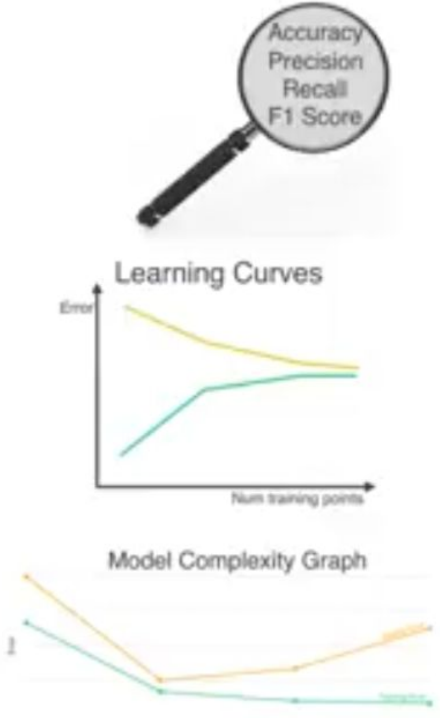
How to use machine learning



Data



Algorithms



Metrics