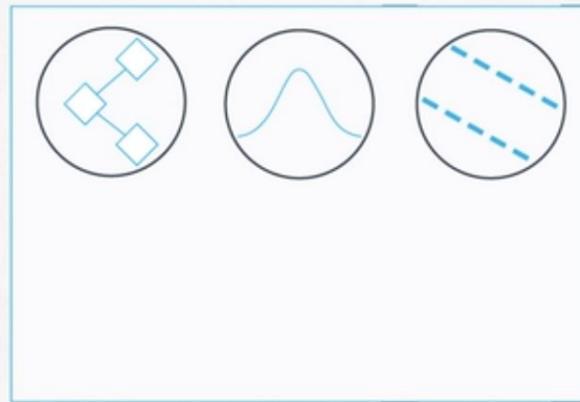
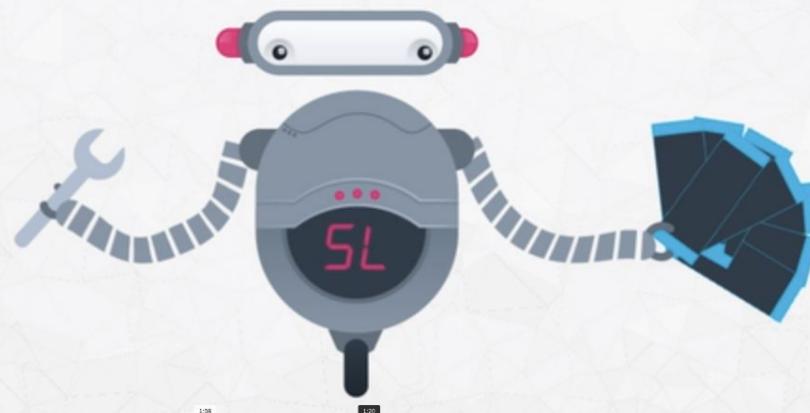
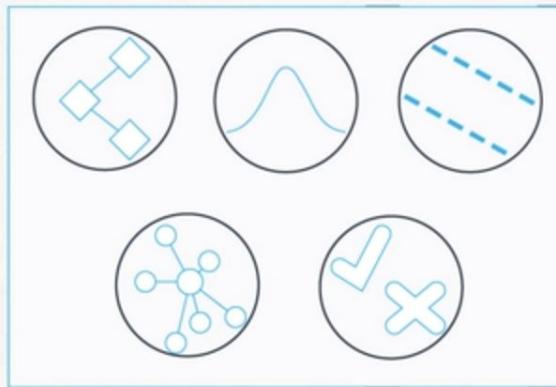
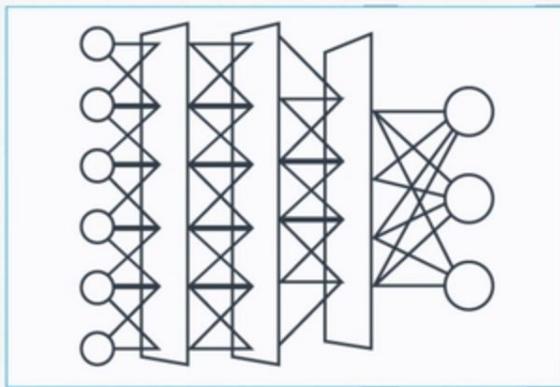


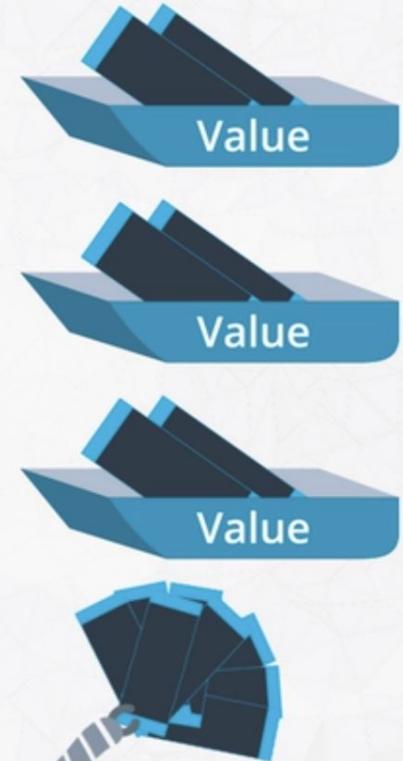
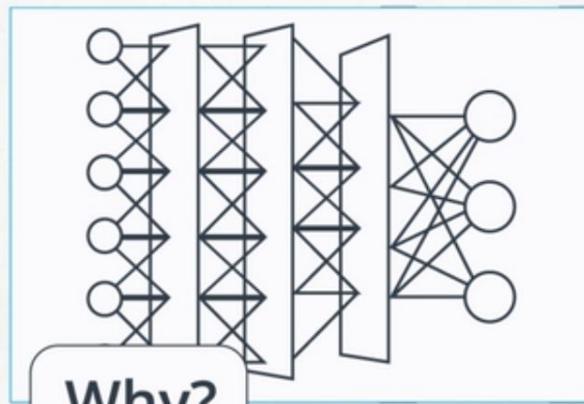
?











Types of Machine Learning

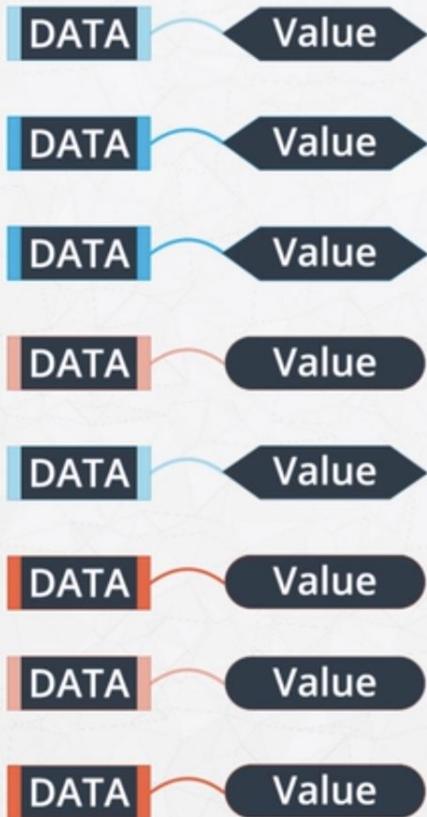
Supervised



Branches of Machine Learning

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Supervised Learning



Supervised Learning

Classification

Categorical Outcomes



Regression

Numeric Outcomes



Unsupervised Learning

DATA

DATA

DATA

DATA

DATA

DATA

DATA

DATA



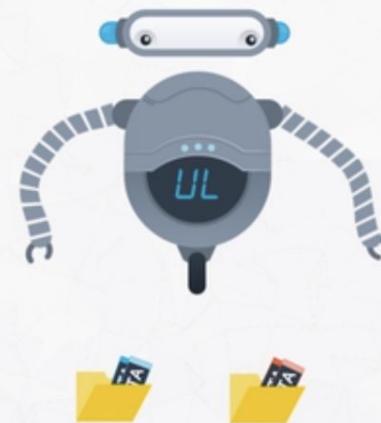
Reinforcement Learning



Supervised Learning



Unsupervised Learning

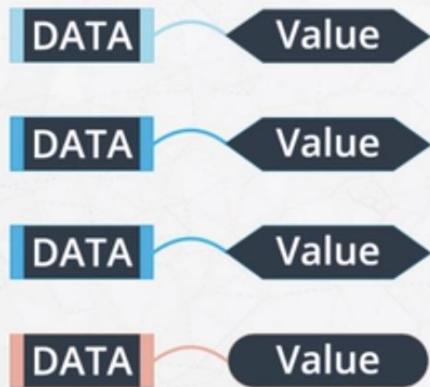


Reinforcement Learning

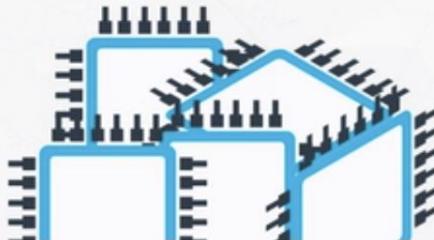


Deep Learning and Neural Networks

Deep Learning

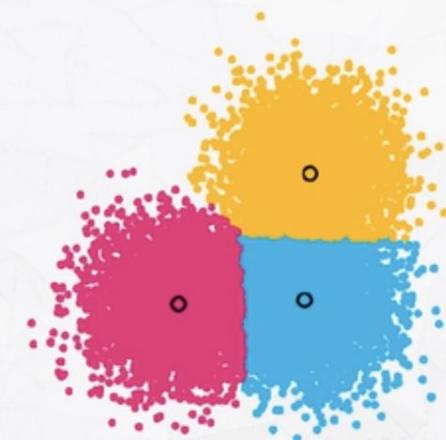
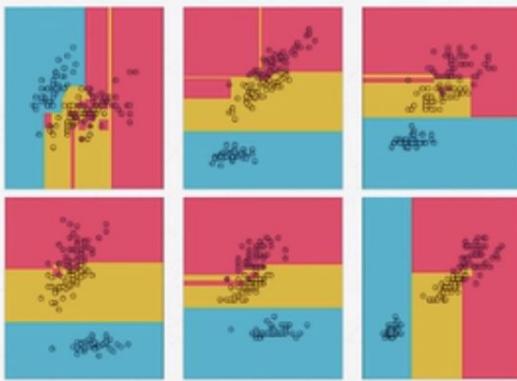


Barriers to Deep Learning

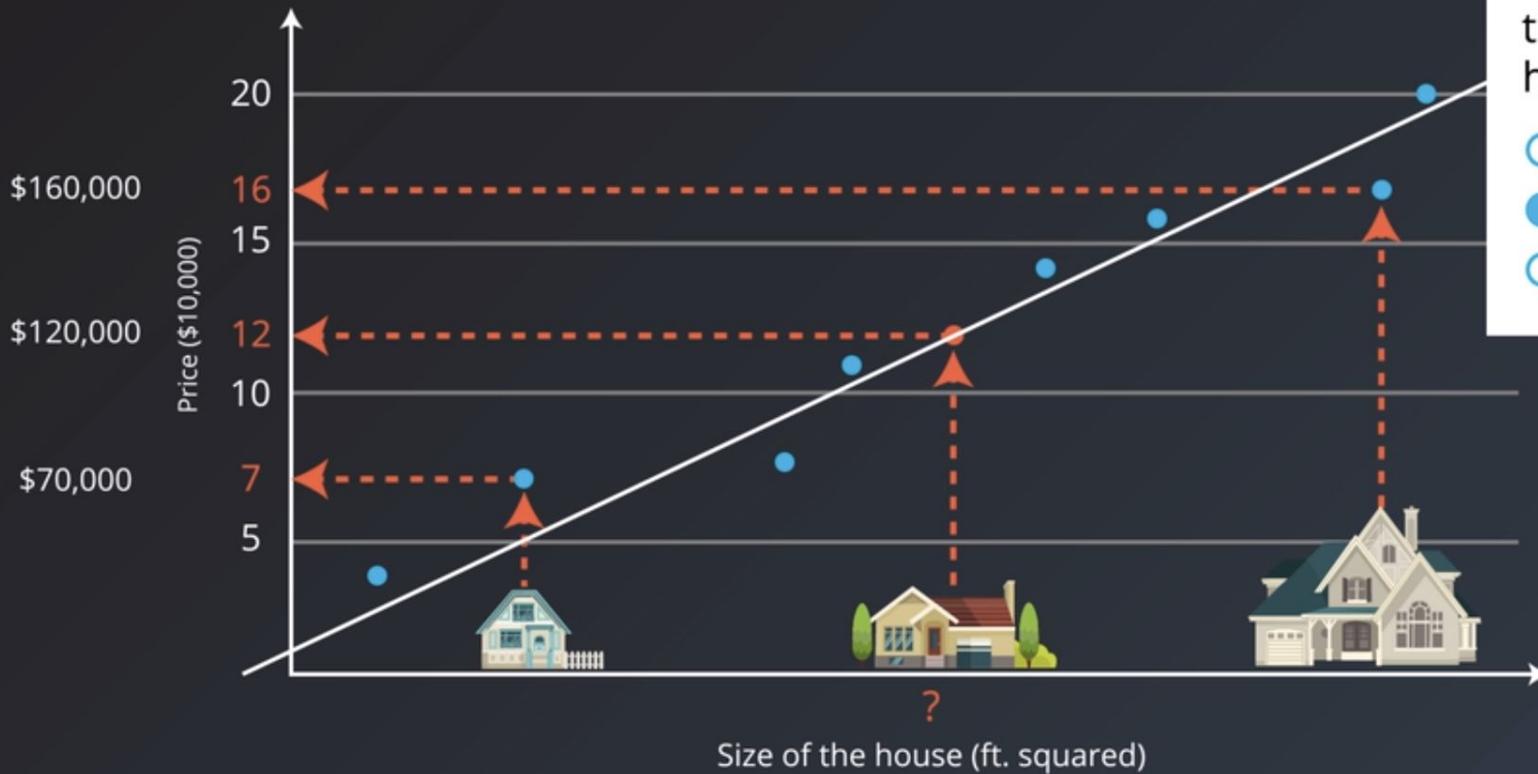




scikit-learn



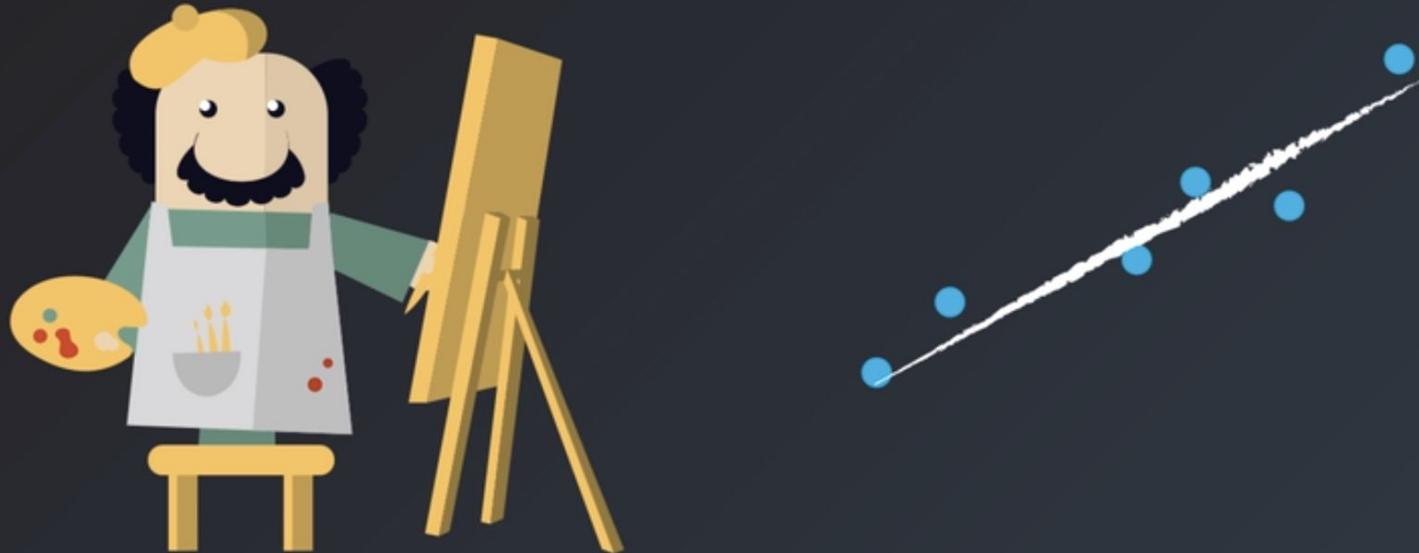
Price of a House



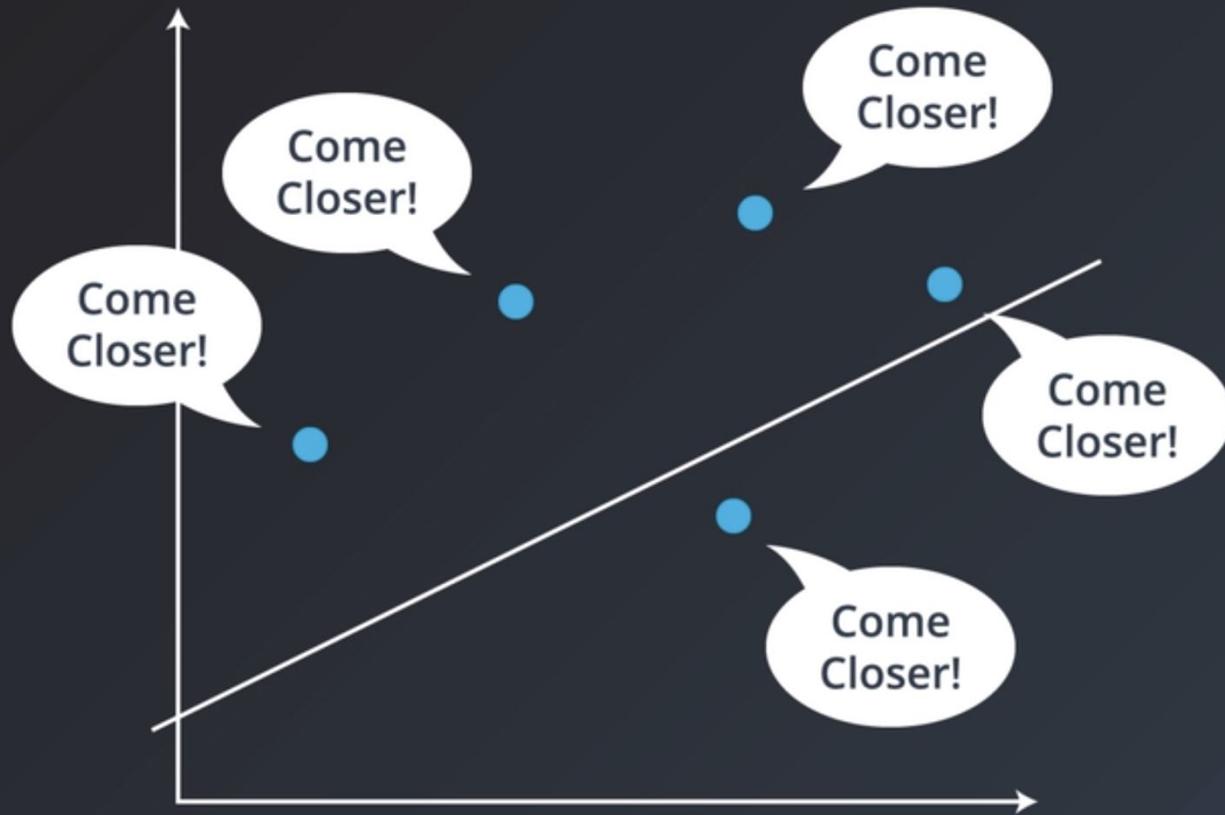
Quiz: What's the best estimate for the price of the house?

- \$80,000
- \$120,000
- \$190,000

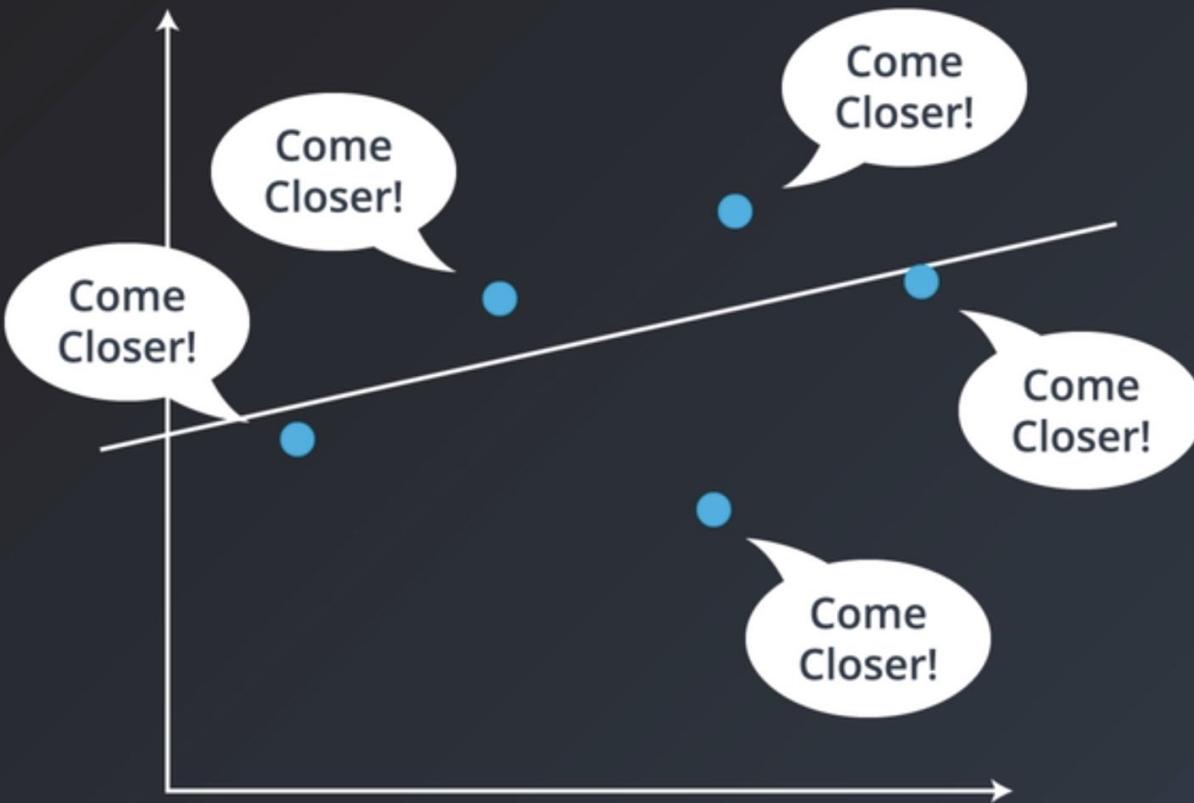
Linear Regression



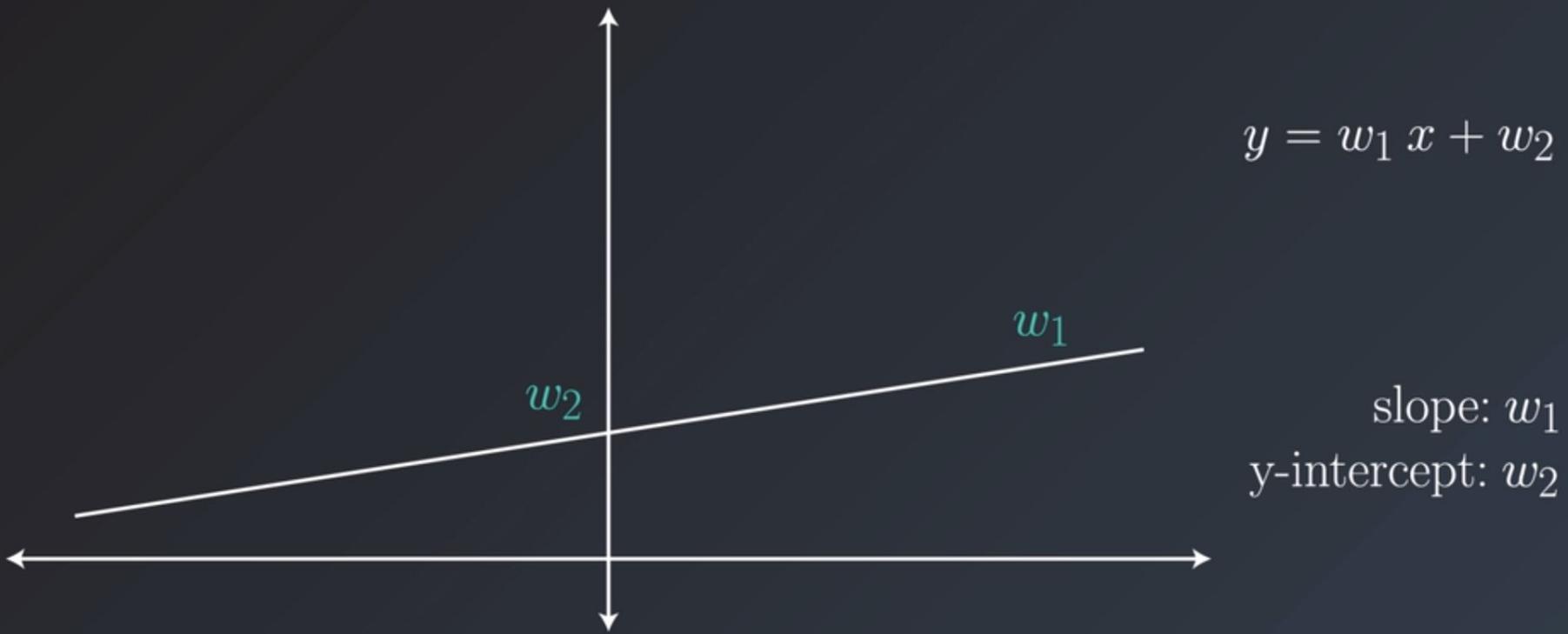
Fitting A Line



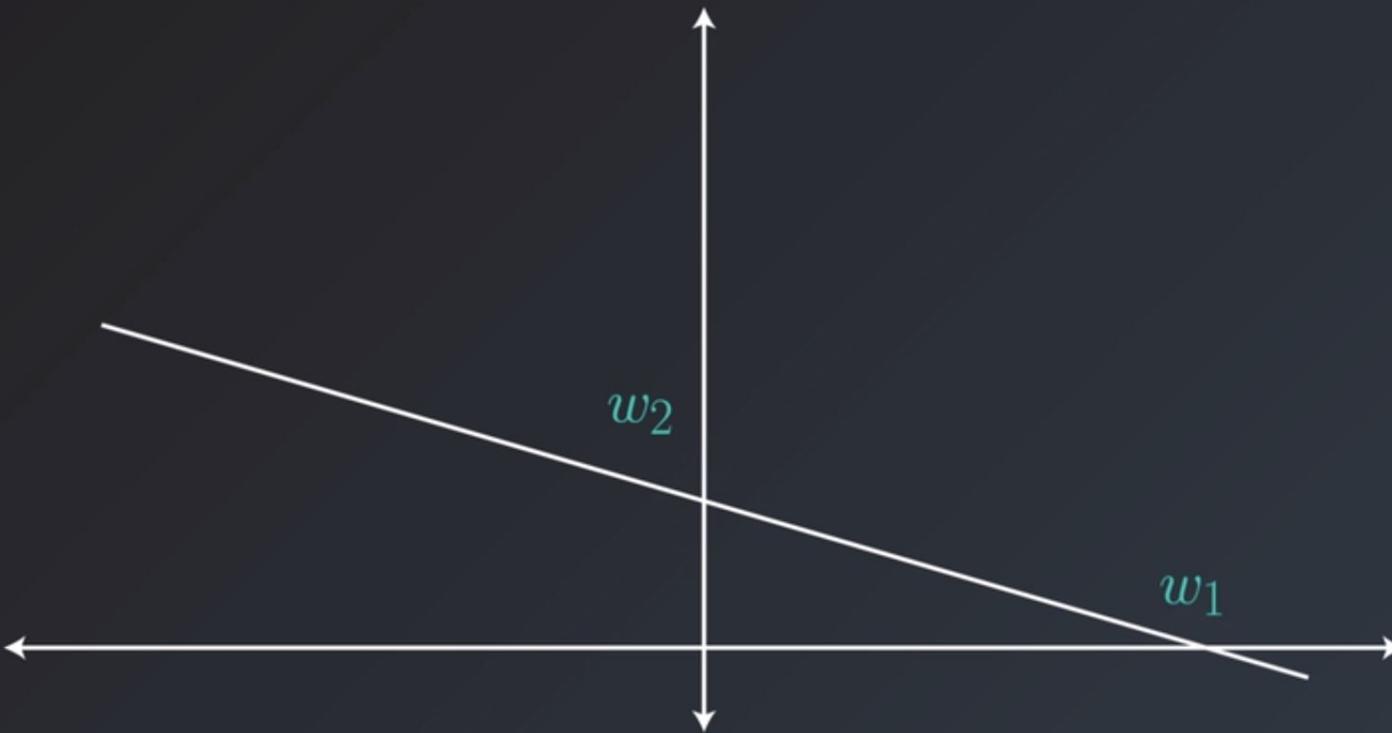
Fitting A Line



Moving A Line



Moving A Line



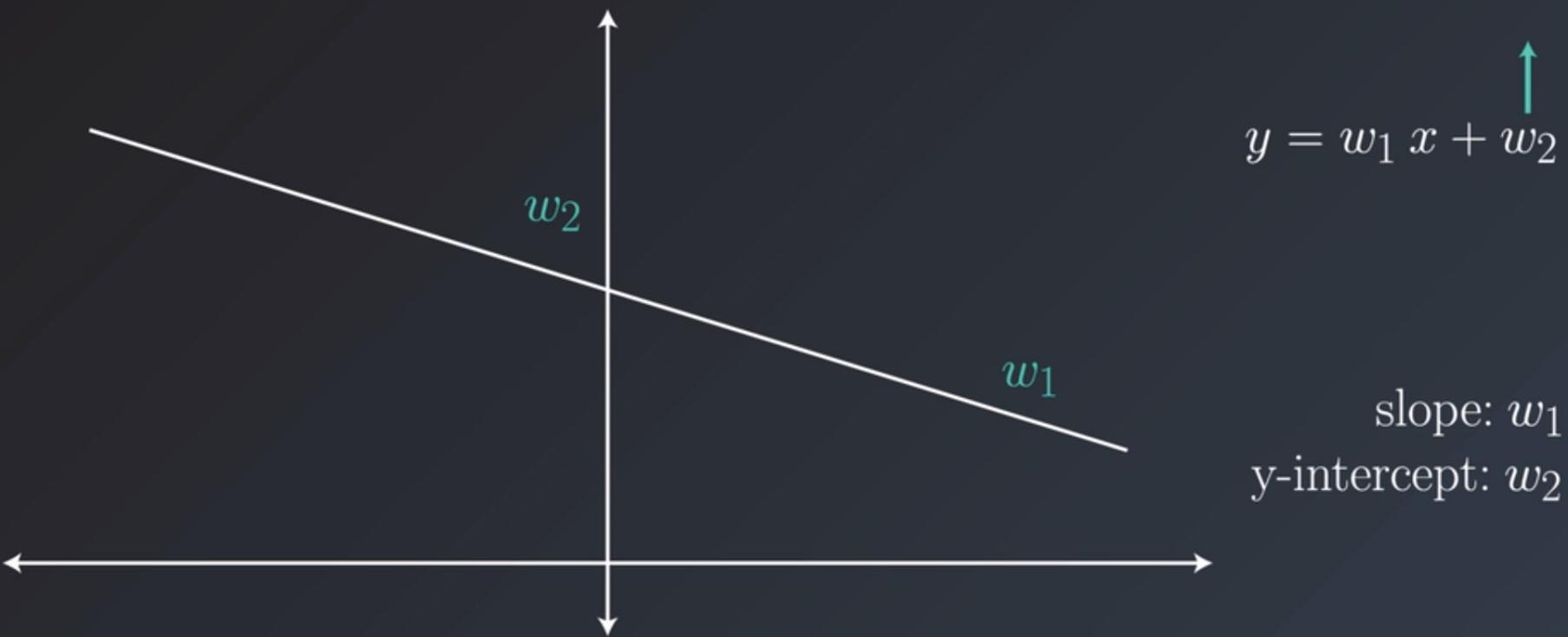
$$y = w_1 x + w_2$$



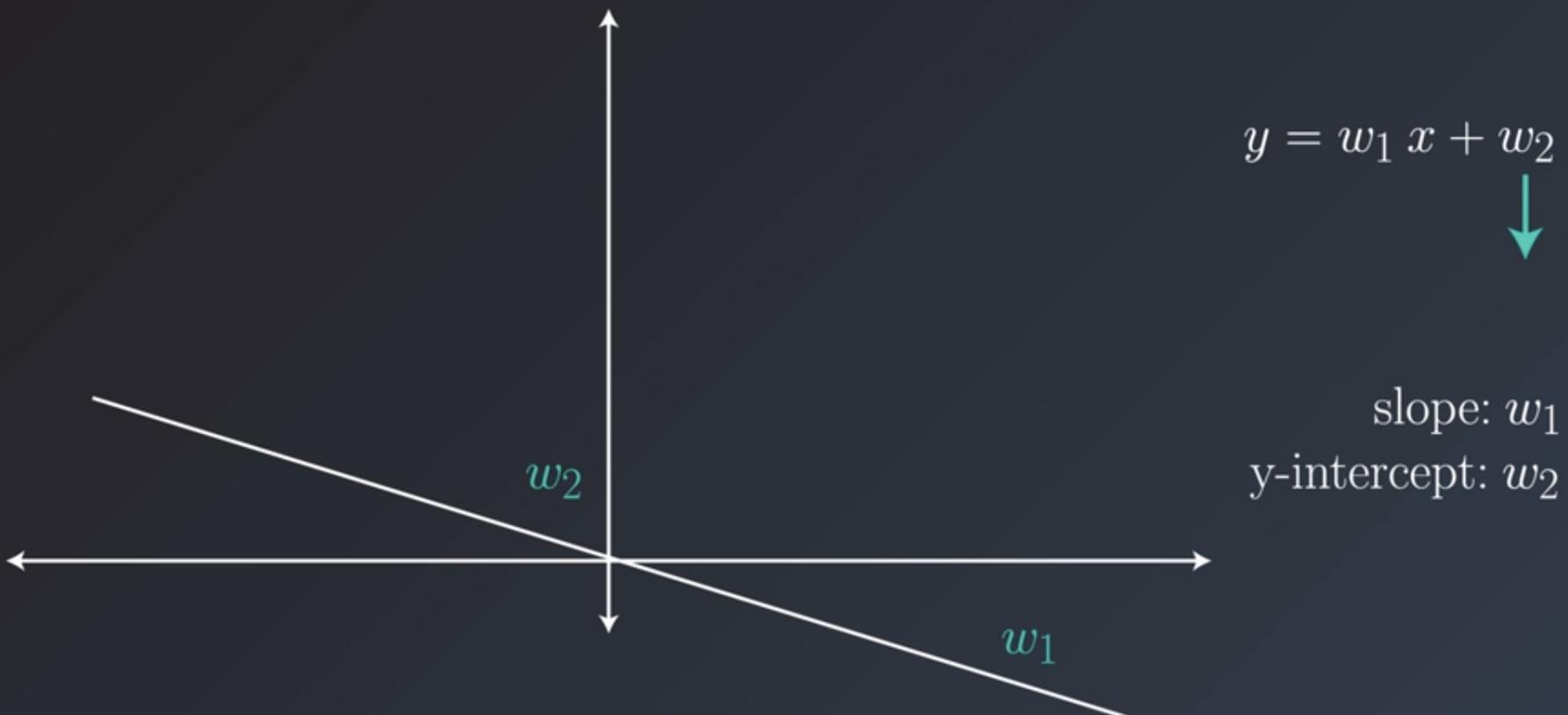
slope: w_1

y-intercept: w_2

Moving A Line



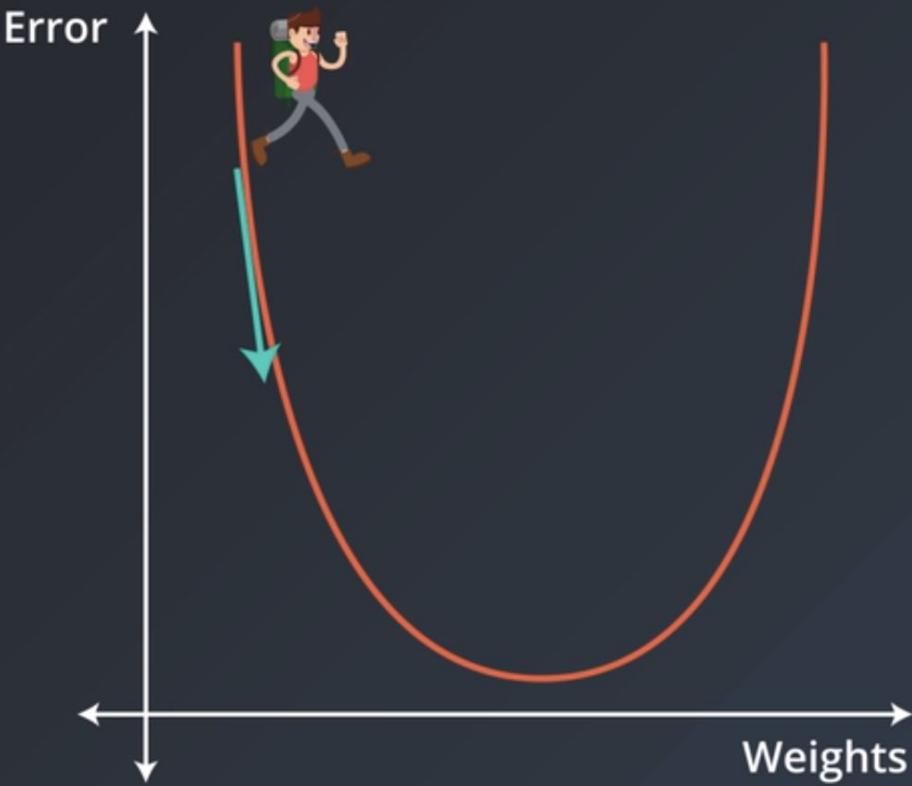
Moving A Line



Gradient Descent

Error Function

- Gradient of
Error Function



Gradient Descent

Error Function

- Gradient of
Error Function

$$w_i \rightarrow w_i - \alpha \frac{\partial}{\partial w_i} Error$$

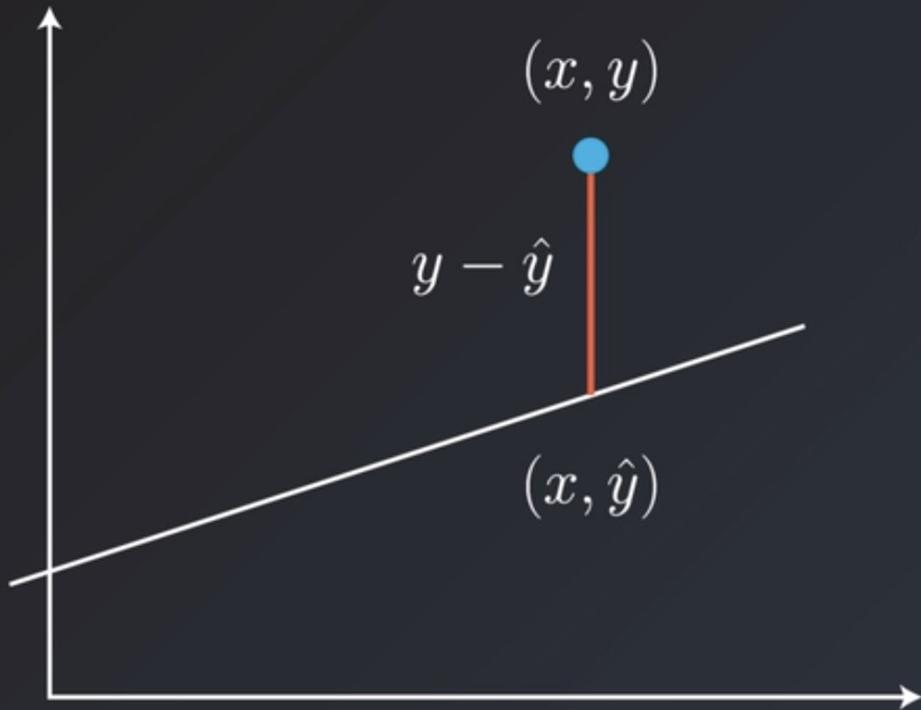


Error Functions

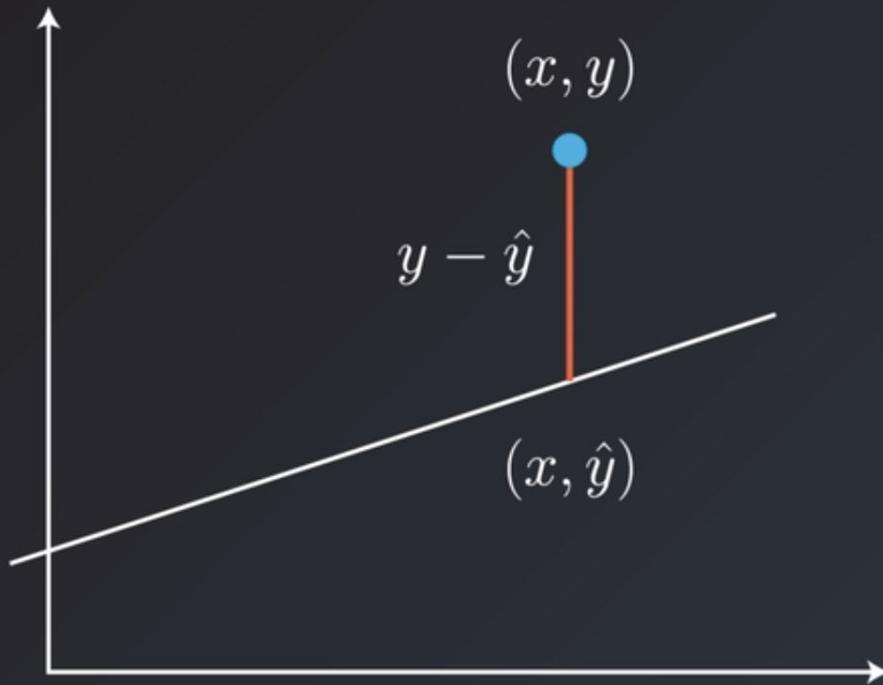
Mean Absolute Error

Mean Squared Error

Mean Absolute Error

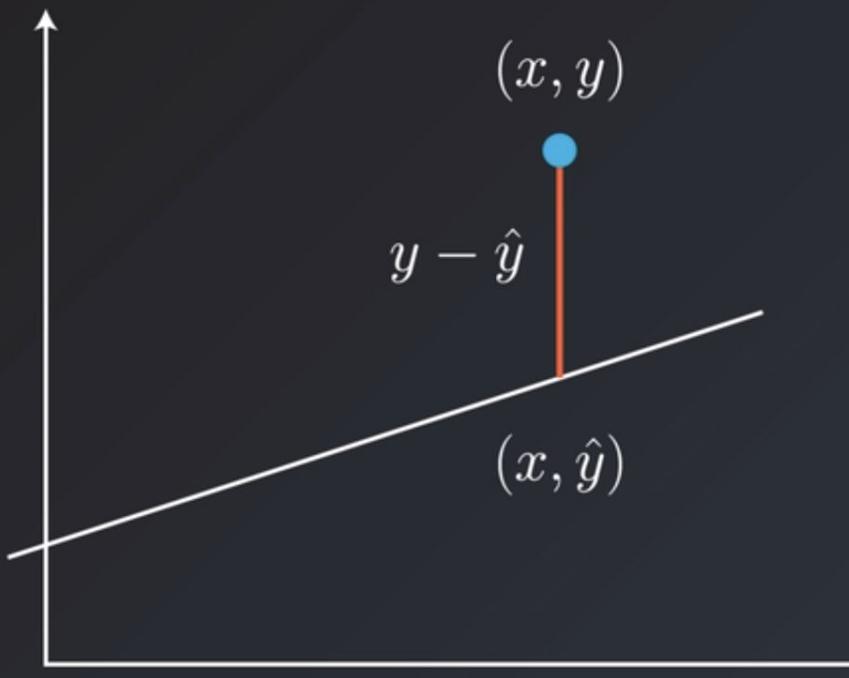


Mean Absolute Error



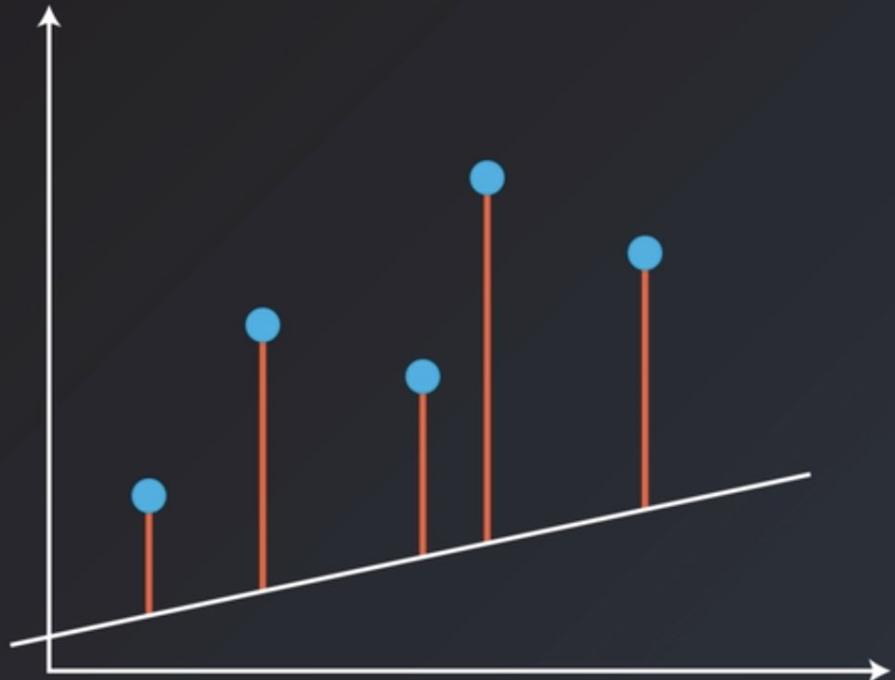
$$Error = \sum_{i=1}^m |y - \hat{y}|$$

Mean Absolute Error

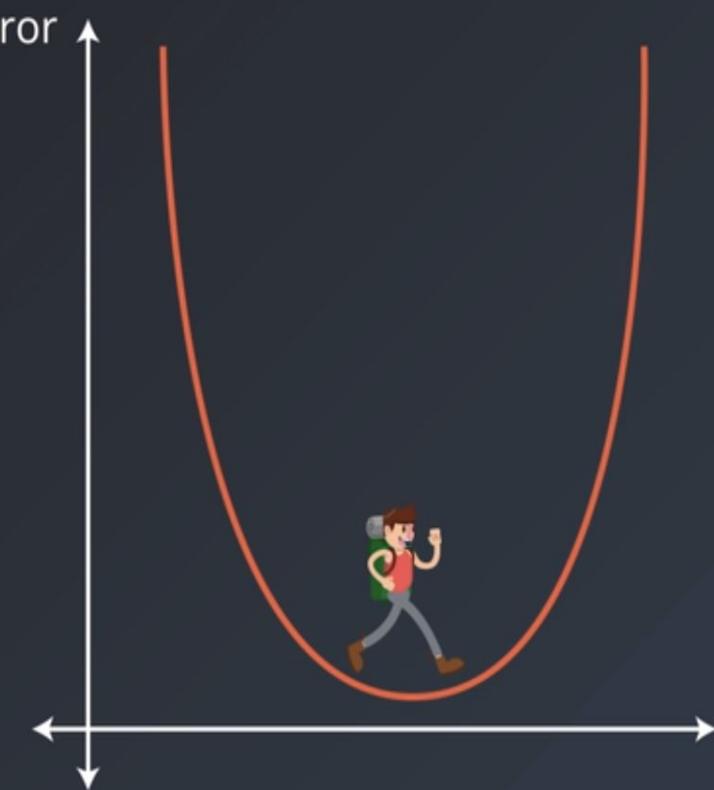
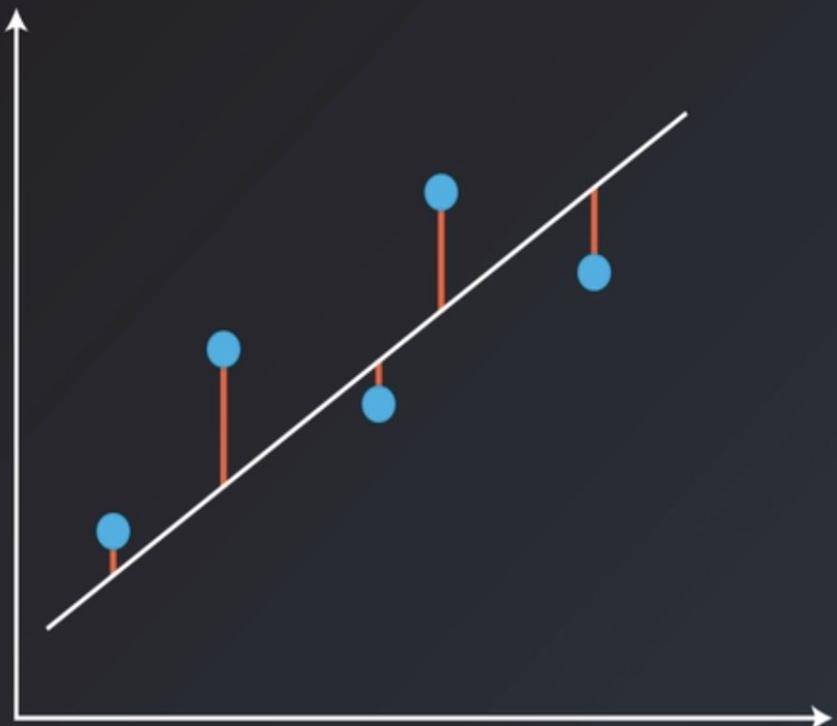


$$Error = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|$$

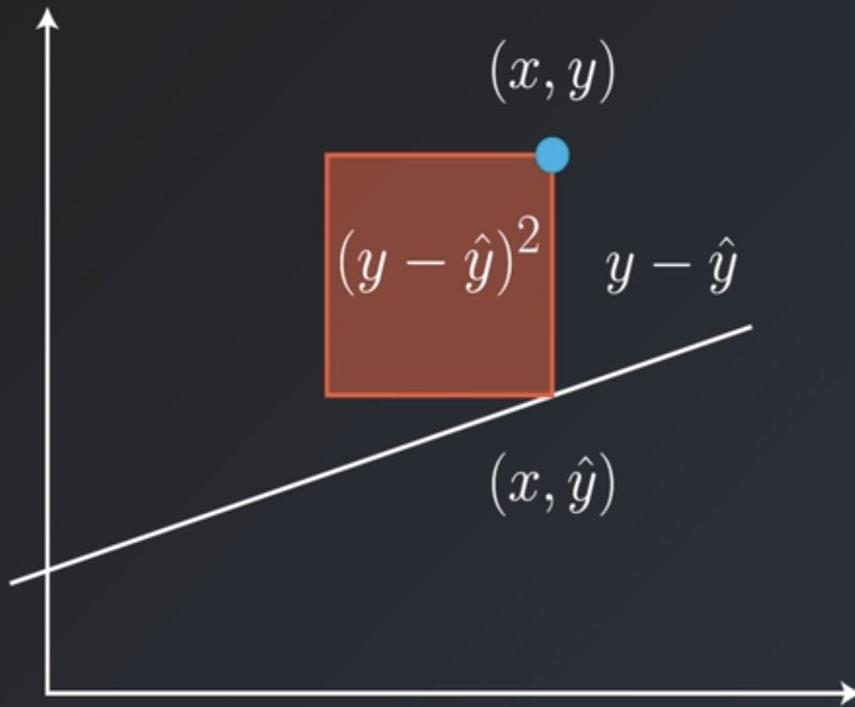
Mean Absolute Error



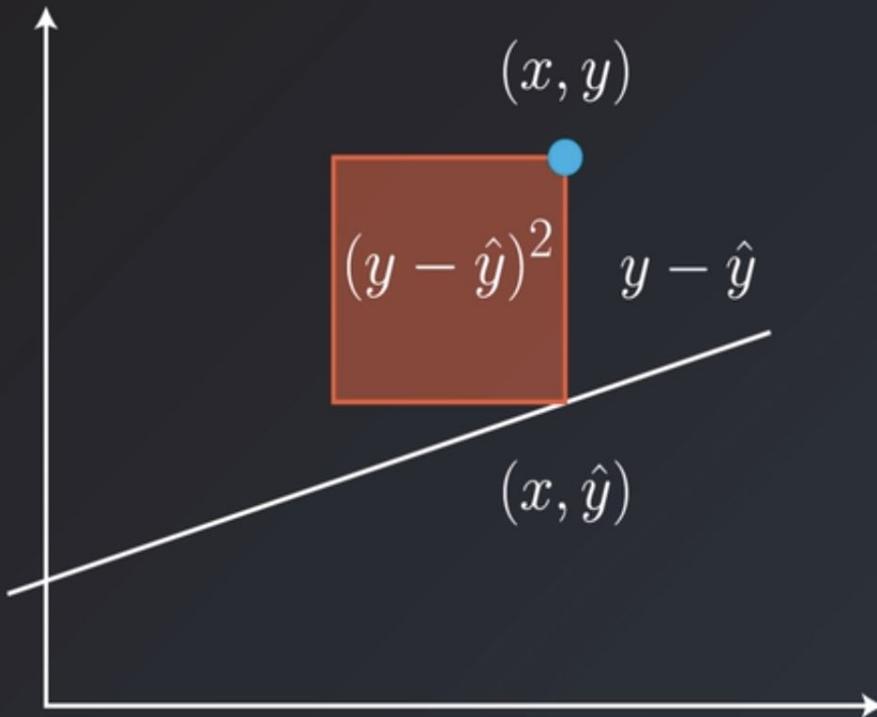
Mean Absolute Error



Mean Squared Error

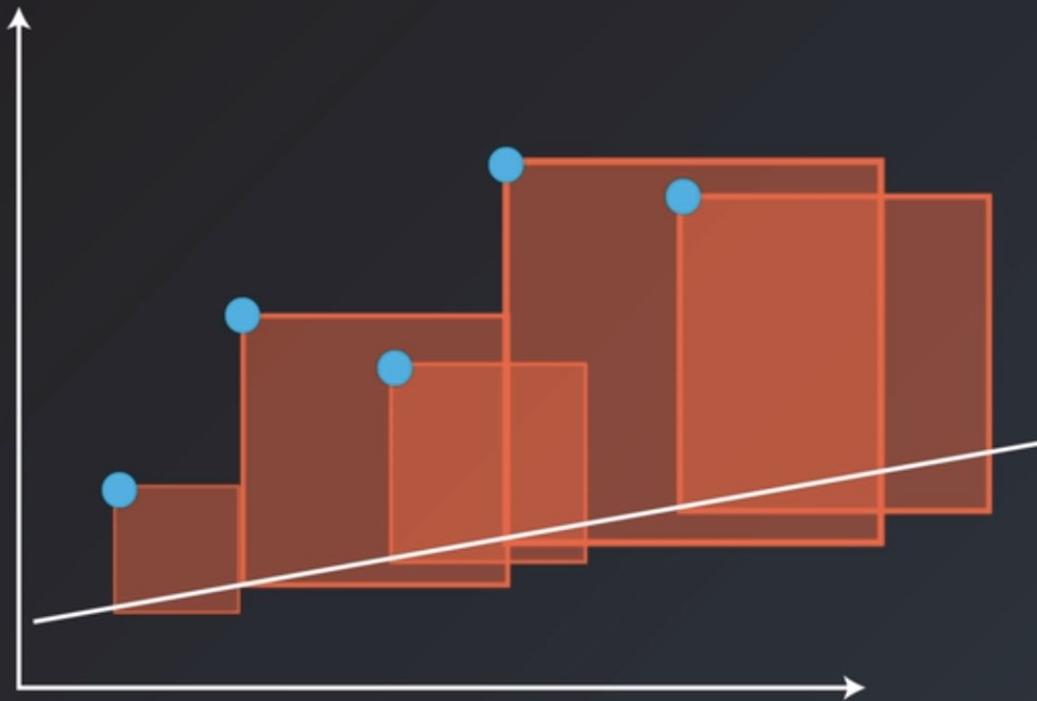


Mean Squared Error



$$Error = \frac{1}{2m} \sum_{i=1}^m (y - \hat{y})^2$$

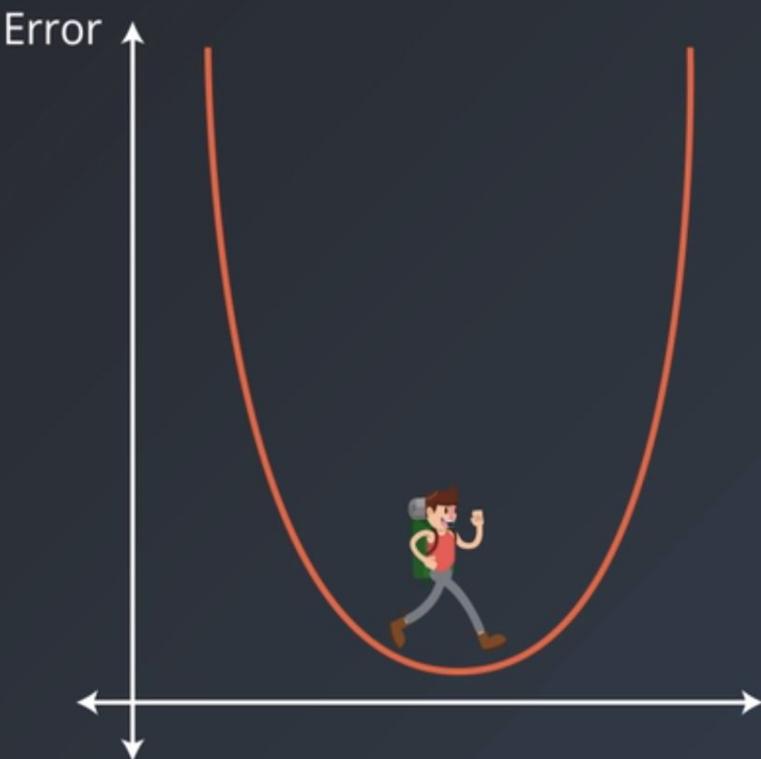
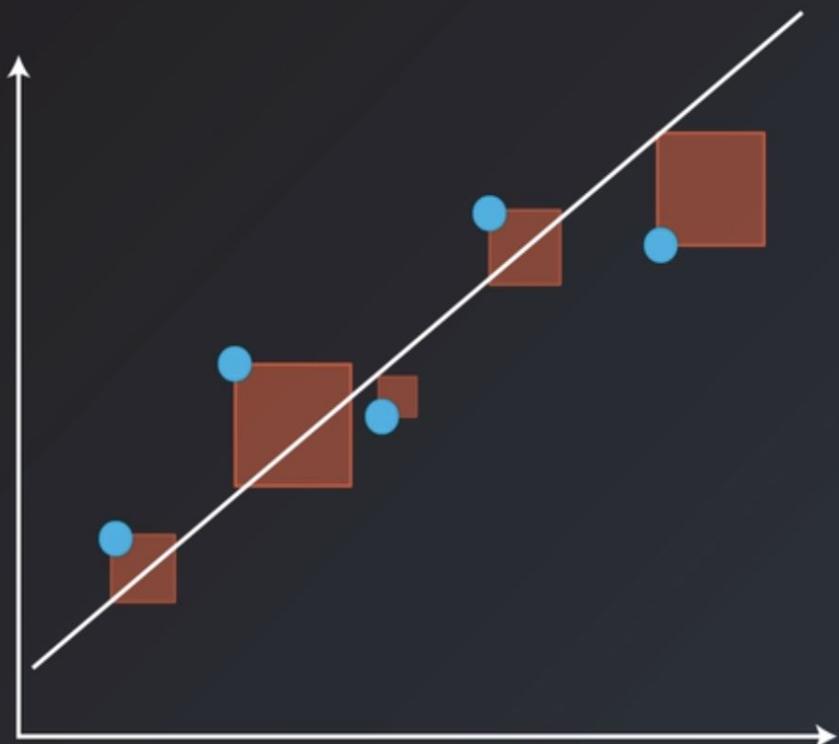
Mean Squared Error



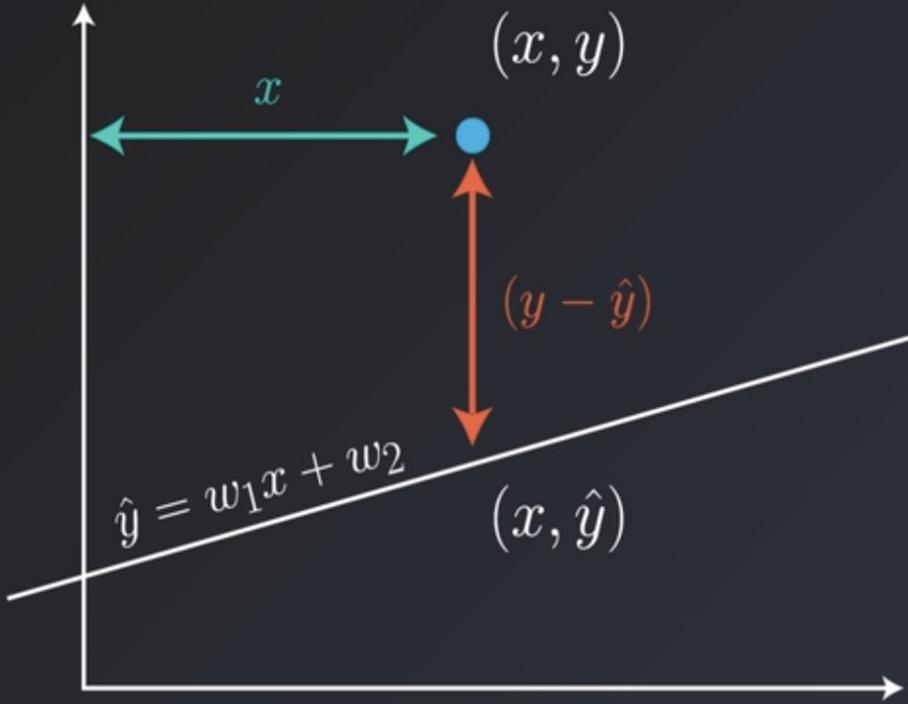
Mean Squared Error



Mean Squared Error



Mean Squared Error



$$\text{Error} = \frac{1}{2}(y - \hat{y})^2$$

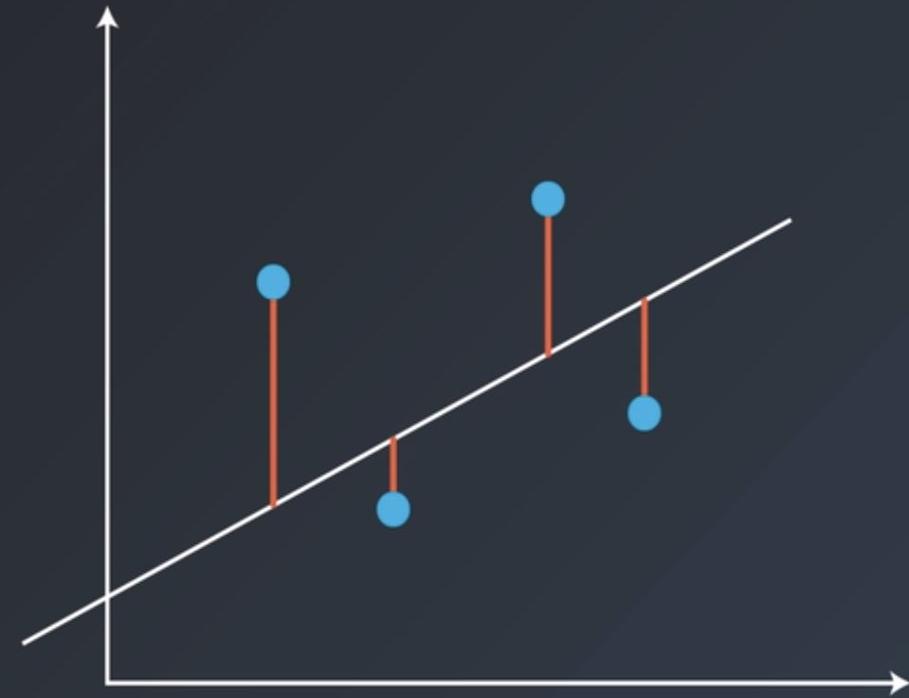
$$\frac{\partial}{\partial w_1} \text{Error} = -(y - \hat{y}) x$$

$$\frac{\partial}{\partial w_2} \text{Error} = -(y - \hat{y})$$

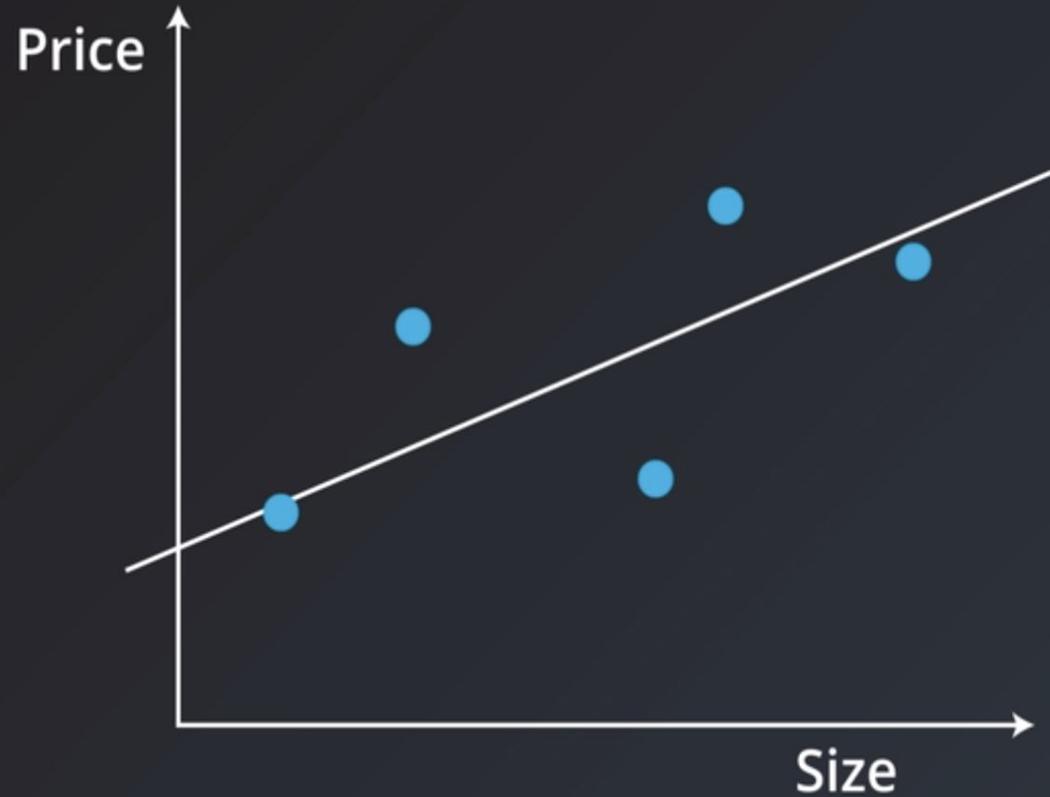
Gradient Step

$$w_i \rightarrow w_i - \alpha \frac{\partial}{\partial w_i} \text{Error}$$

Mean Absolute Error



Housing Prices



Prediction: A line

$$\text{Price} = w_1(\text{size}) + w_2$$

Housing Prices



Prediction: A plane

$$\text{Price} = w_1(\text{school quality}) + w_2(\text{size}) + w_3$$

Housing Prices

	Size	School quality	...	Num. Rooms	Price
House 1	900	6	...	2	\$100k
House 2	560	4	...	1	\$50k
...	
House m	2000	7	...	4	\$250k

Housing Prices

	x_1	x_2	\dots	x_{n-1}	\hat{y}
	Size	School quality	...	Num. Rooms	Price
House 1	900	6	...	2	\$100k
House 2	560	4	...	1	\$50k
...	
House m	2000	7	...	4	\$250k

↔ n columns ↔

n dimensional space

x_1, x_2, \dots, x_{n-1}

Housing Prices

	x_1	x_2	\dots	x_{n-1}	\hat{y}
	Size	School quality	...	Num. Rooms	Price
House 1	900	6	...	2	\$100k
House 2	560	4	...	1	\$50k
...	
House m	2000	7	...	4	\$250k

↔ n columns ↔

n dimensional space

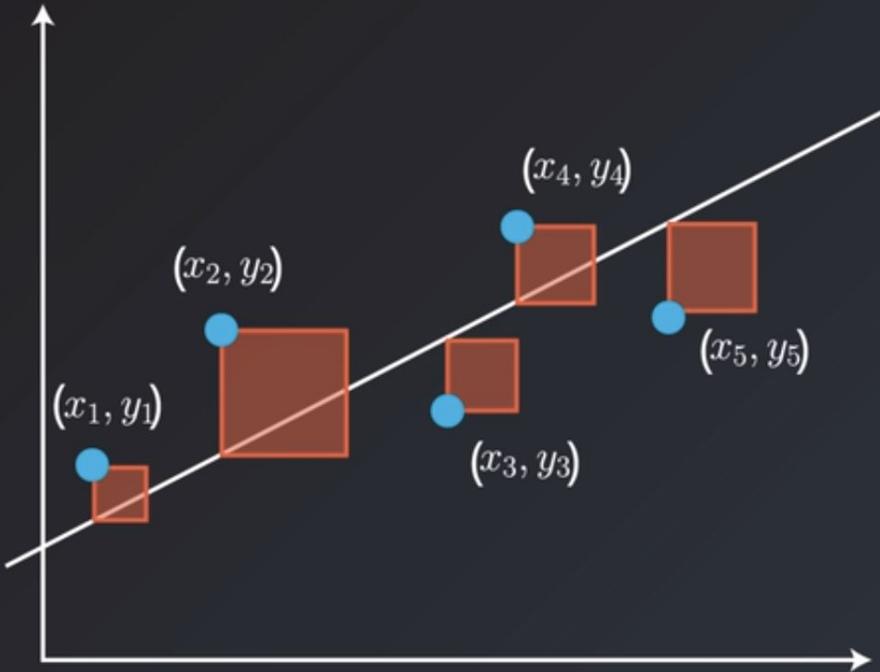
x_1, x_2, \dots, x_{n-1}

Prediction

$n-1$ dimensional hyperplane

$$\hat{y} = w_1x_1 + w_2x_2 + \dots + w_{n-1}x_{n-1} + w_n$$

Mean Squared Error



x_1, x_2, \dots, x_m

y_1, y_2, \dots, y_m

$$\hat{y}_i = w_1 x_i + w_2$$

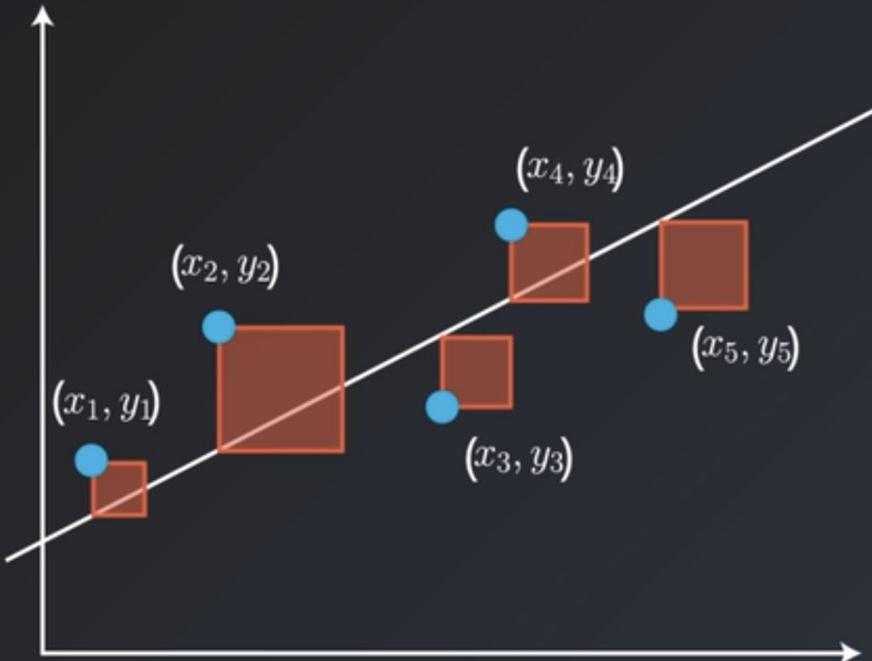
$$E(w_1, w_2) = \frac{1}{2m} \sum_{i=1}^m (\hat{y} - y)^2$$

$$0 = \frac{\sum x_i^2}{m} w_1 + \frac{\sum x_i}{m} w_2 + \frac{\sum y_i}{m}$$

$$0 = \frac{\sum x_i}{m} w_1 + w_2 + \frac{\sum y_i}{m}$$

2 equations
2 unknowns

Mean Squared Error



x_1, x_2, \dots, x_m

y_1, y_2, \dots, y_m

$$\hat{y}_i = w_1 x_i + w_2$$

$$E(w_1, w_2) = \frac{1}{2m} \sum_{i=1}^m (\hat{y} - y)^2$$

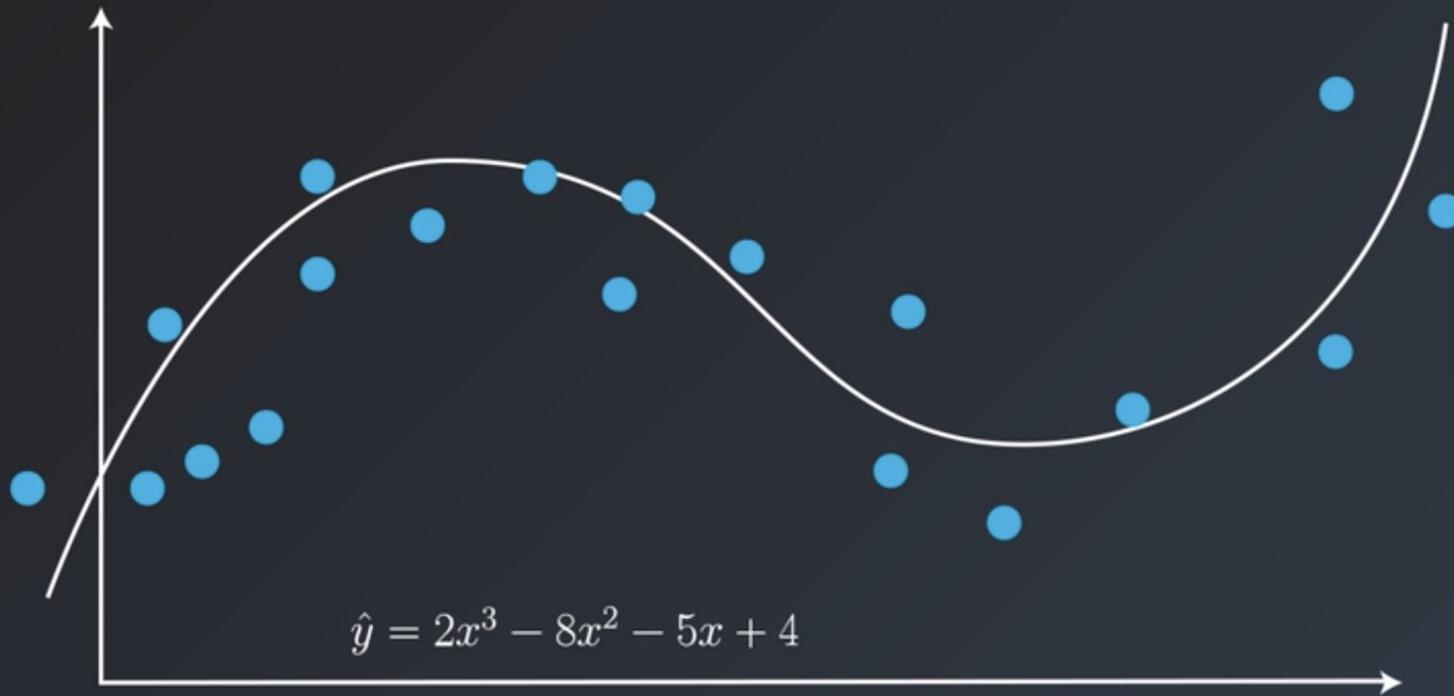
$$\boxed{0 = \frac{\sum x_i^2}{m} w_1 + \frac{\sum x_i}{m} w_2 + \frac{\sum x_i y_i}{m}}$$
$$0 = \frac{\sum x_i}{m} w_1 + w_2 + \frac{\sum y_i}{m}$$

n equations
n unknowns

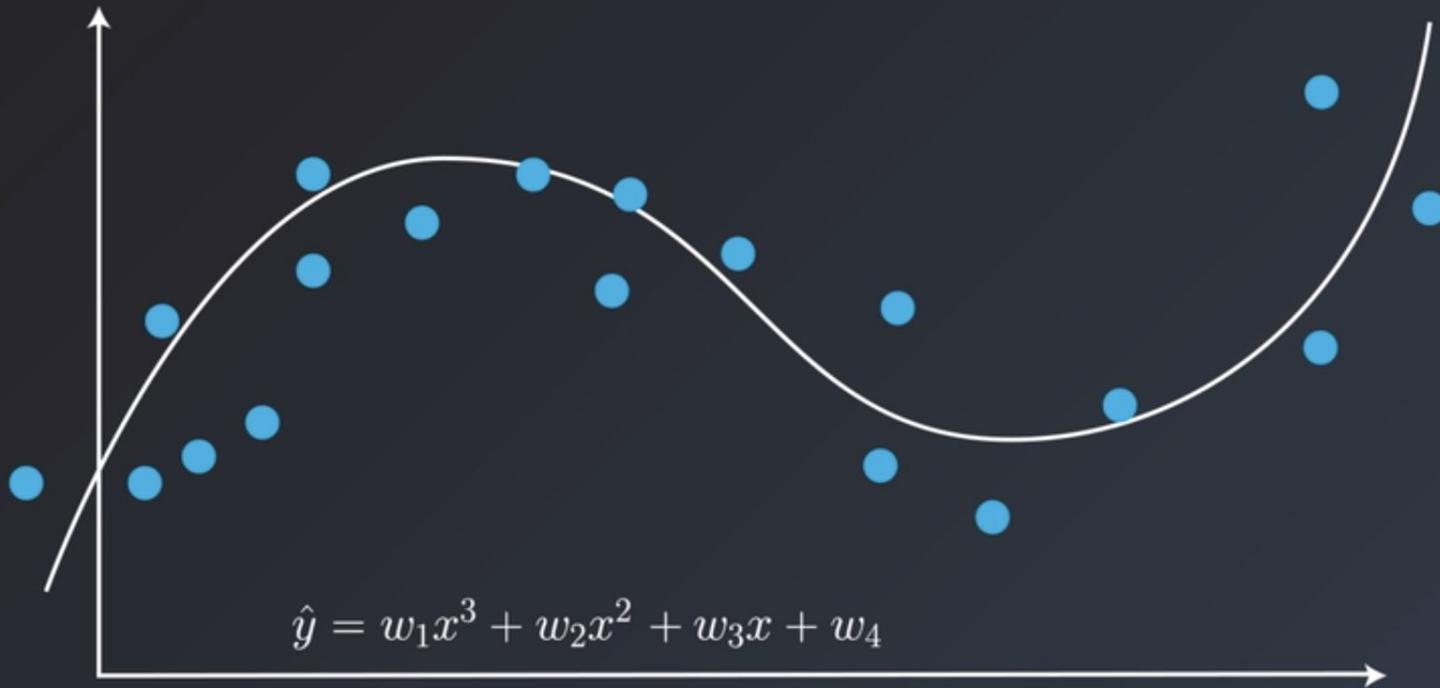
Polynomial Regression



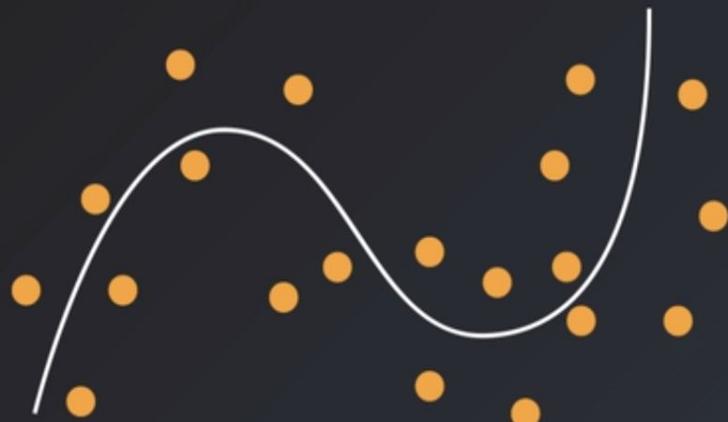
Polynomial Regression



Polynomial Regression



Regression and Classification

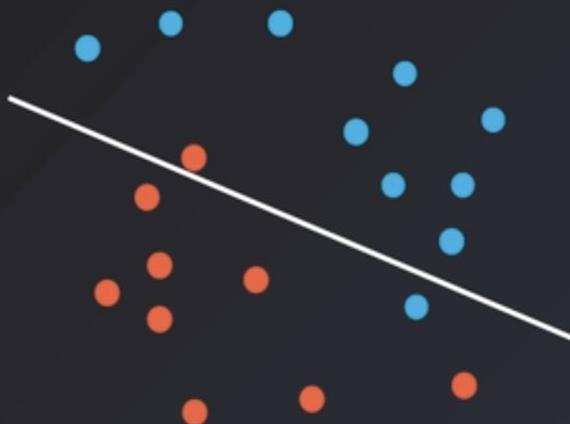


Regression

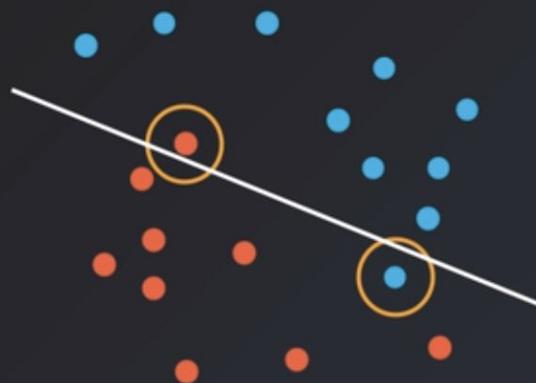


Classification

Different Models



Different Models



Error



$$3x_1 + 4x_2 + 5 = 0$$



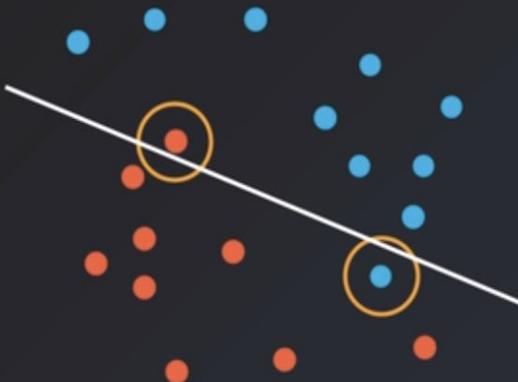
Error



$$2x_1^3 - 2x_1^2x_2 - 4x_2^3 + 3x_1^2 + 6x_1x_2 + 4x_2^2 + 5 = 0$$



Different Models



Error



$$3x_1 + 4x_2 + 5 = 0$$

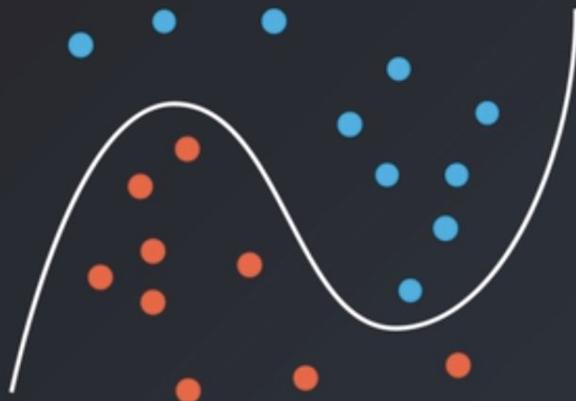


Error



$$2x_1^3 - 2x_1^2x_2 - 4x_2^3 + 3x_1^2 + 6x_1x_2 + 4x_2^2 + 5 = 0$$

L1 Regularization



$$2x_1^3 - 2x_1^2x_2 - 4x_2^3 + 3x_1^2 + 6x_1x_2 + 4x_2^2 + 5 = 0$$

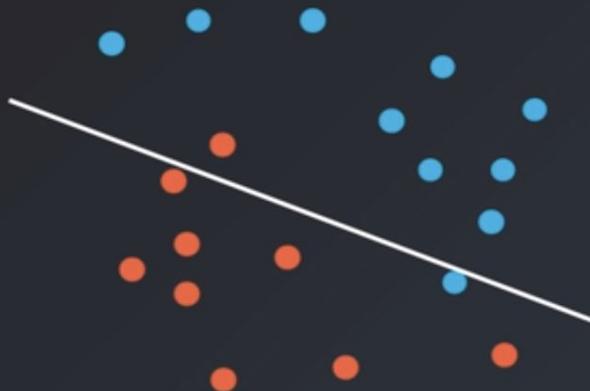
L1 Regularization



$$2x_1^3 - 2x_1^2x_2 - 4x_2^3 + 3x_1^2 + 6x_1x_2 + 4x_2^2 + 5 = 0$$

$$\text{Error} = |2| + |-2| + |-4| + |3| + |6| + |4| = 21$$

L1 Regularization



$$3x_1 + 4x_2 + 5 = 0$$

$$\text{Error} = |3| + |4| = 7$$

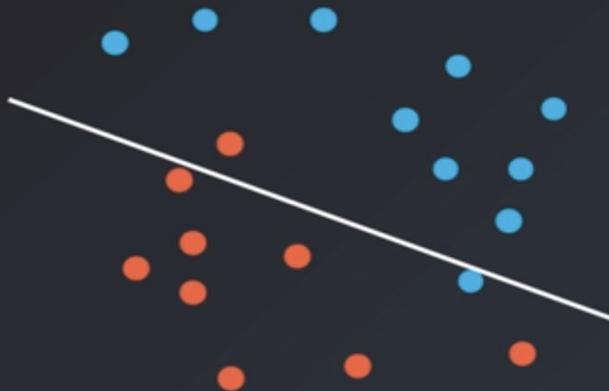
L2 Regularization



$$2x_1^3 - 2x_1^2x_2 - 4x_2^3 + 3x_1^2 + 6x_1x_2 + 4x_2^2 + 5 = 0$$

$$\text{Error} = 2^2 + (-2)^2 + (-4)^2 + 3^2 + 6^2 + 4^2 = 85$$

L2 Regularization



$$3x_1 + 4x_2 + 5 = 0$$

$$\text{Error} = 3^2 + 4^2 = 25$$

Simple vs Complex Models



Requires low error
OK if it's complex
Punishment on the complexity should be small

Requires simplicity
OK with errors
Punishment on the complexity should be large

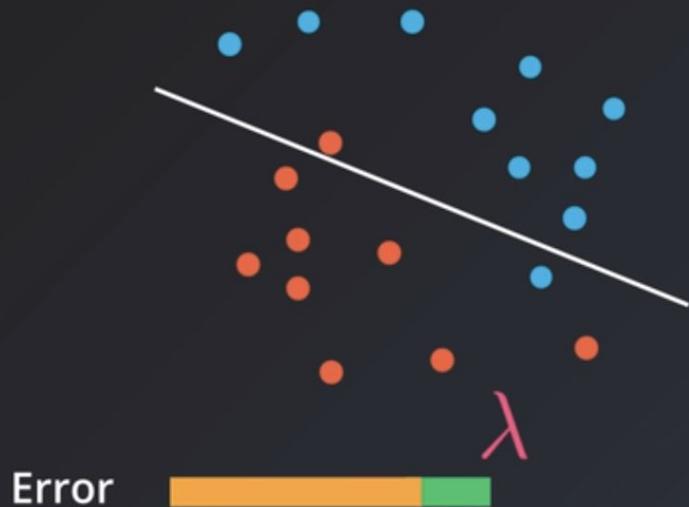
The λ Parameter



$$3x_1 + 4x_2 + 5 = 0$$

$$2x_1^3 - 2x_1^2x_2 - 4x_2^3 + 3x_1^2 + 6x_1x_2 + 4x_2^2 + 5 = 0$$

The λ Parameter



Small λ

Error

$$3x_1 + 4x_2 + 5 = 0$$

$$2x_1^3 - 2x_1^2x_2 - 4x_2^3 + 3x_1^2 + 6x_1x_2 + 4x_2^2 + 5 = 0$$



The λ Parameter



$$3x_1 + 4x_2 + 5 = 0$$

$$2x_1^3 - 2x_1^2x_2 - 4x_2^3 + 3x_1^2 + 6x_1x_2 + 4x_2^2 + 5 = 0$$

Simple vs Complex Models



Requires low error
OK if it's complex

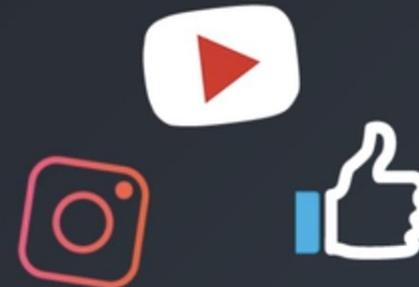


Requires simplicity
OK with errors

Simple vs Complex Models



Requires low error
OK if it's complex
Punishment on the complexity should be small



Requires simplicity
OK with errors
Punishment on the complexity should be large

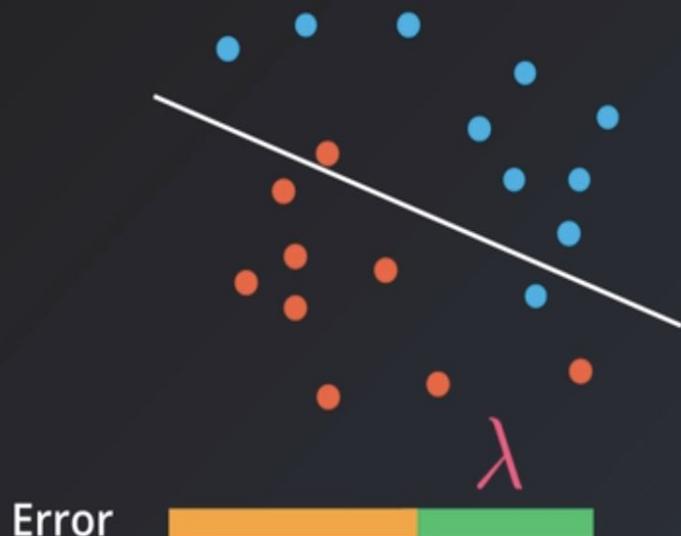
The λ Parameter



$$3x_1 + 4x_2 + 5 = 0$$

$$2x_1^3 - 2x_1^2x_2 - 4x_2^3 + 3x_1^2 + 6x_1x_2 + 4x_2^2 + 5 = 0$$

The λ Parameter



$$3x_1 + 4x_2 + 5 = 0$$



$$2x_1^3 - 2x_1^2x_2 - 4x_2^3 + 3x_1^2 + 6x_1x_2 + 4x_2^2 + 5 = 0$$

The λ Parameter



$$3x_1 + 4x_2 + 5 = 0$$

$$2x_1^3 - 2x_1^2x_2 - 4x_2^3 + 3x_1^2 + 6x_1x_2 + 4x_2^2 + 5 = 0$$

L1 vs L2 Regularization

L1 Regularization	L2 Regularization
Computationally Inefficient (unless data is sparse)	Computationally Efficient
Sparse Outputs	Non-Sparse Outputs
Feature Selection	No Feature Selection