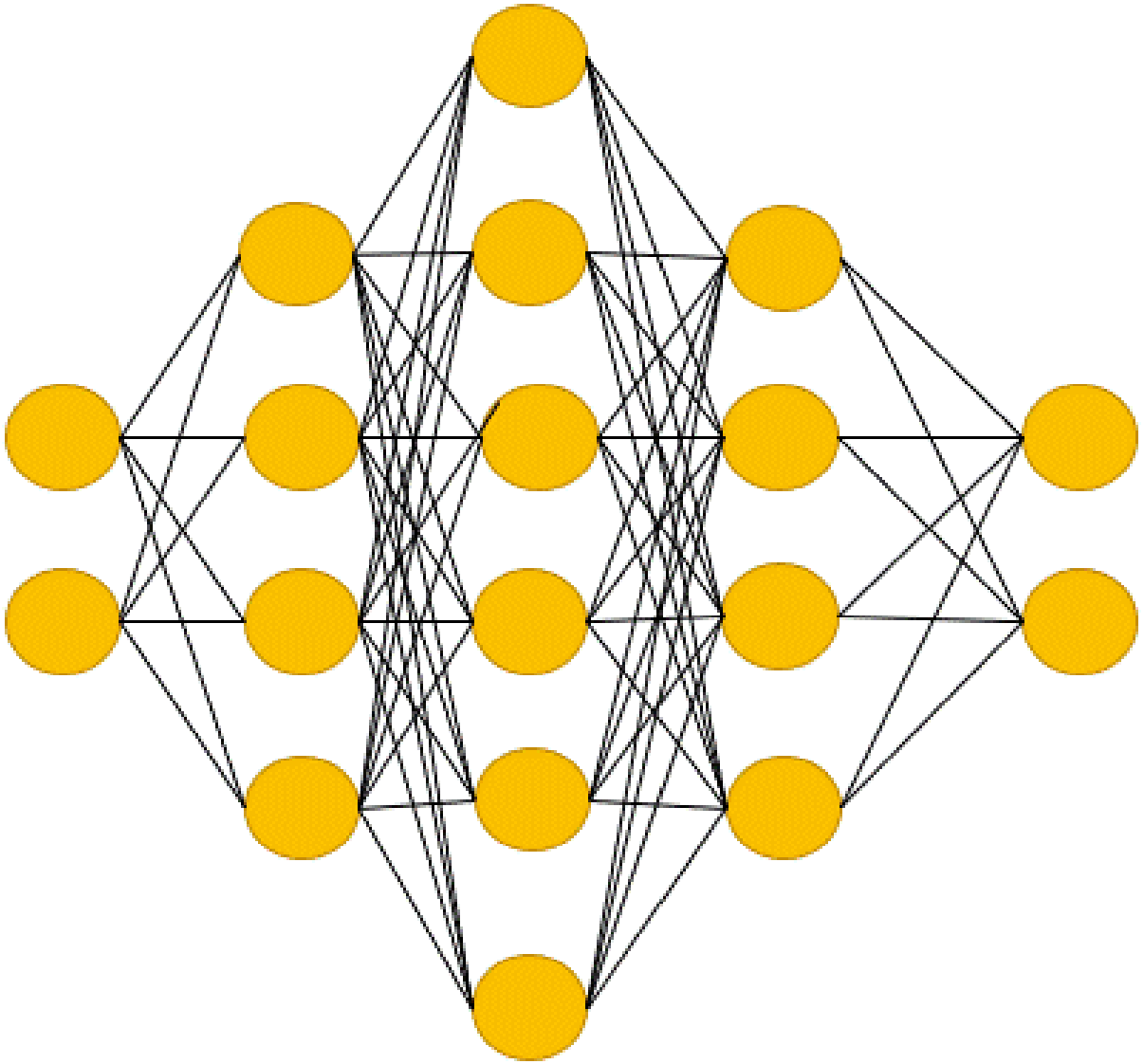
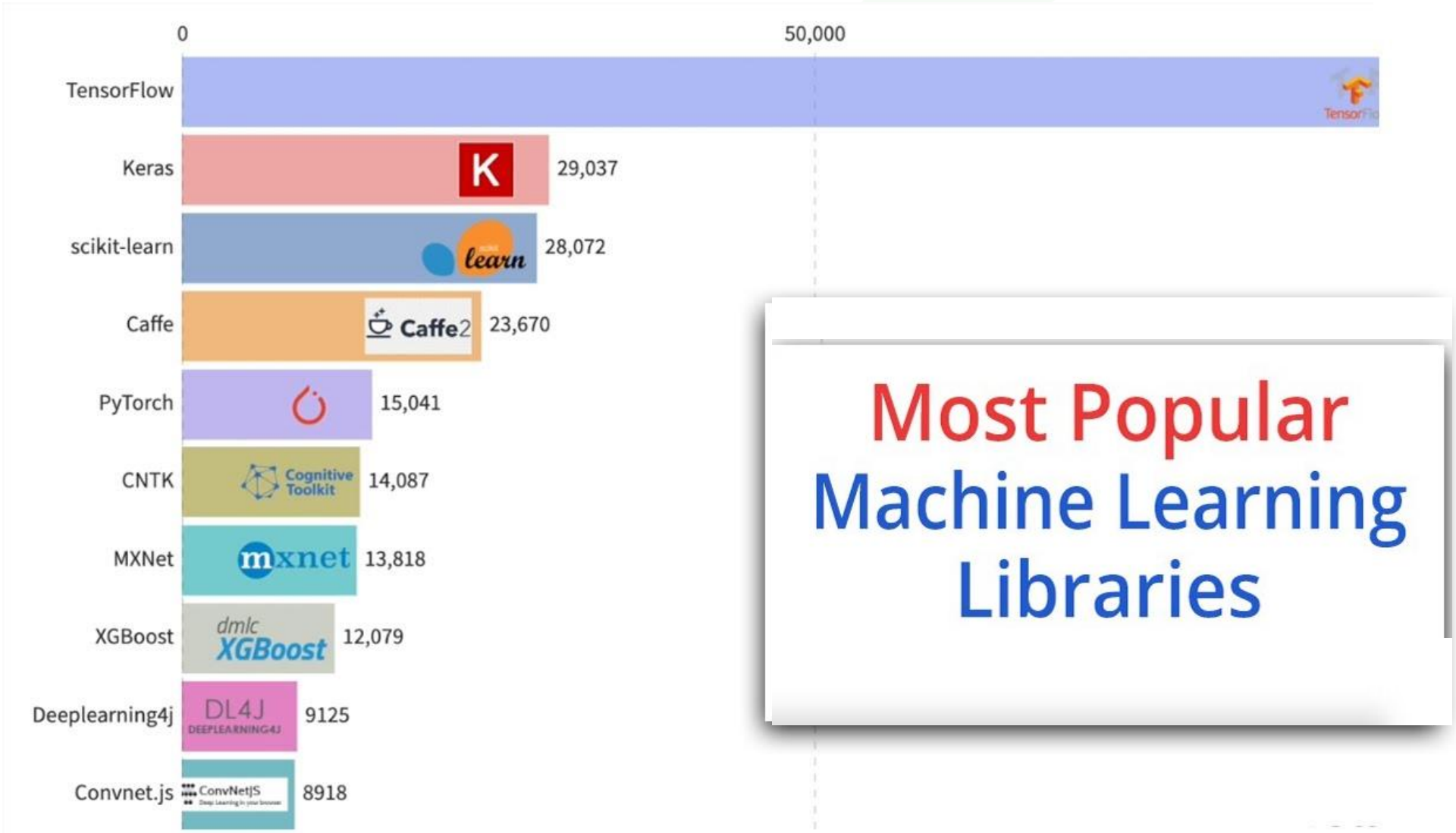
Let me ask, Gru

What is the Price of the Jet, Bob? It must be 50K



MOST POPULAR DEEP LEARNING LIBRARIES&PLATFORMS

DEEP LEARNING KÜTÜPHANELERİ



DEEP LEARNING KÜTÜPHANELERİ

Keras

Keras is an open source neural network library written in Python. It is capable of running on top of TensorFlow. It is designed to enable fast experimentation with **deep neural networks**.

TensorFlow

TensorFlow is an open-source software library for dataflow programming across a range of tasks. It is a symbolic math library that is used for **machine learning** applications like neural networks.

PyTorch

PyTorch is an open source **machine learning** library for Python, based on Torch. It is used for applications such as **natural language processing** and was developed by Facebook's AI research group.

Theano

Caffe2

Caffe is a deep learning framework, originally developed at University of California, Berkeley. It is open source, under a BSD license. It is written in C++, with a Python interface.

DEEP LEARNING KÜTÜPHANELERİ

	Languages	Tutorials and training materials	CNN modeling capability	RNN modeling capability	Architecture: easy-to-use and modular front end	Speed	Multiple GPU support	Keras compatible
Theano	Python, C++	++	++	++	+	++	+	+
Tensor-Flow	Python	+++	+++	++	+++	++	++	+
Torch	Lua, Python (new)	+	+++	++	++	+++	++	
Caffe	C++	+	++		+	+	+	
MXNet	R, Python, Julia, Scala	++	++	+	++	++	+++	
Neon	Python	+	++	+	+	++	+	
CNTK	C++	+	+	+++	+	++	+	

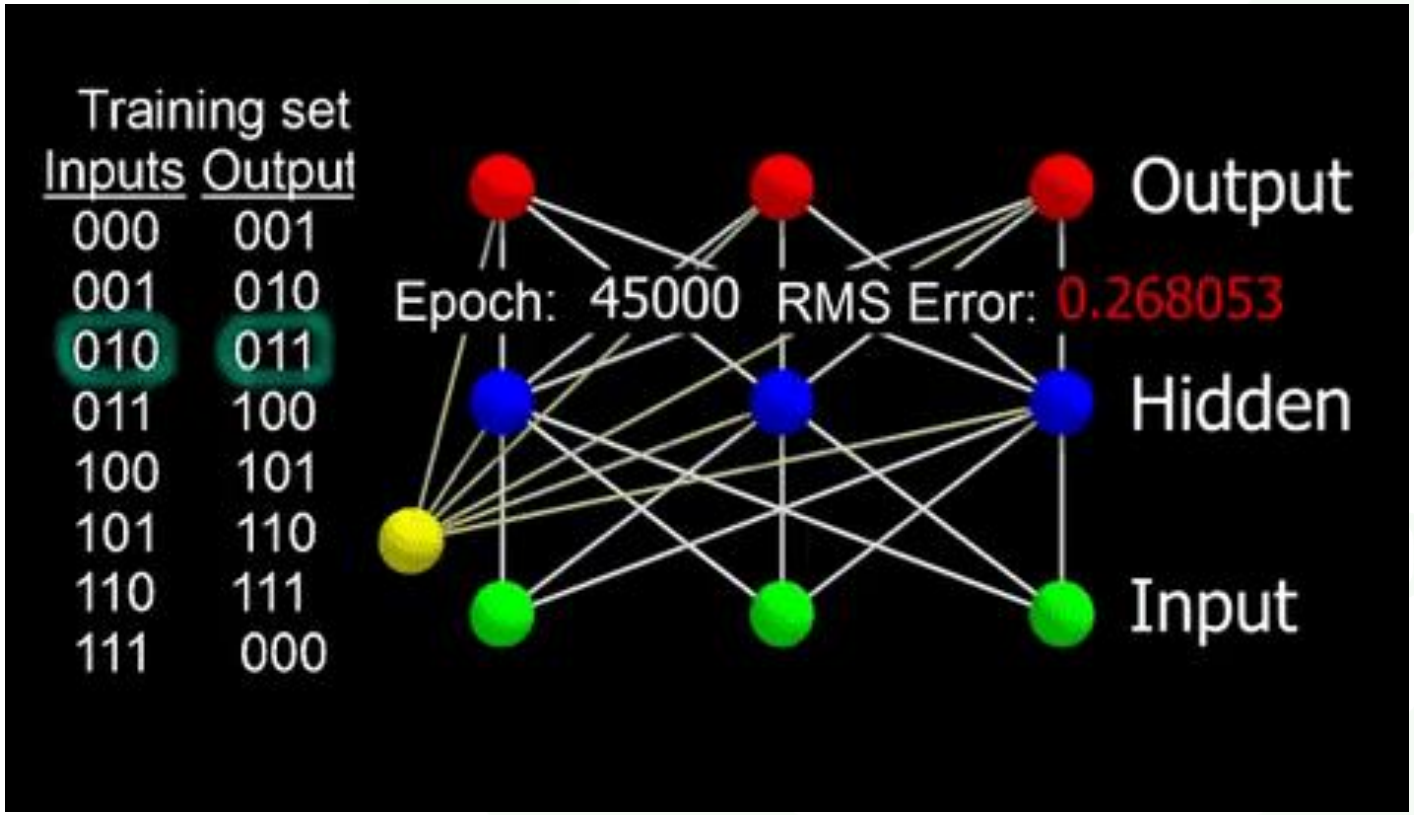
EPOCH, BATCHSIZE


ER
EFUL
BELT
TION.
KITCHEN

EPOCH

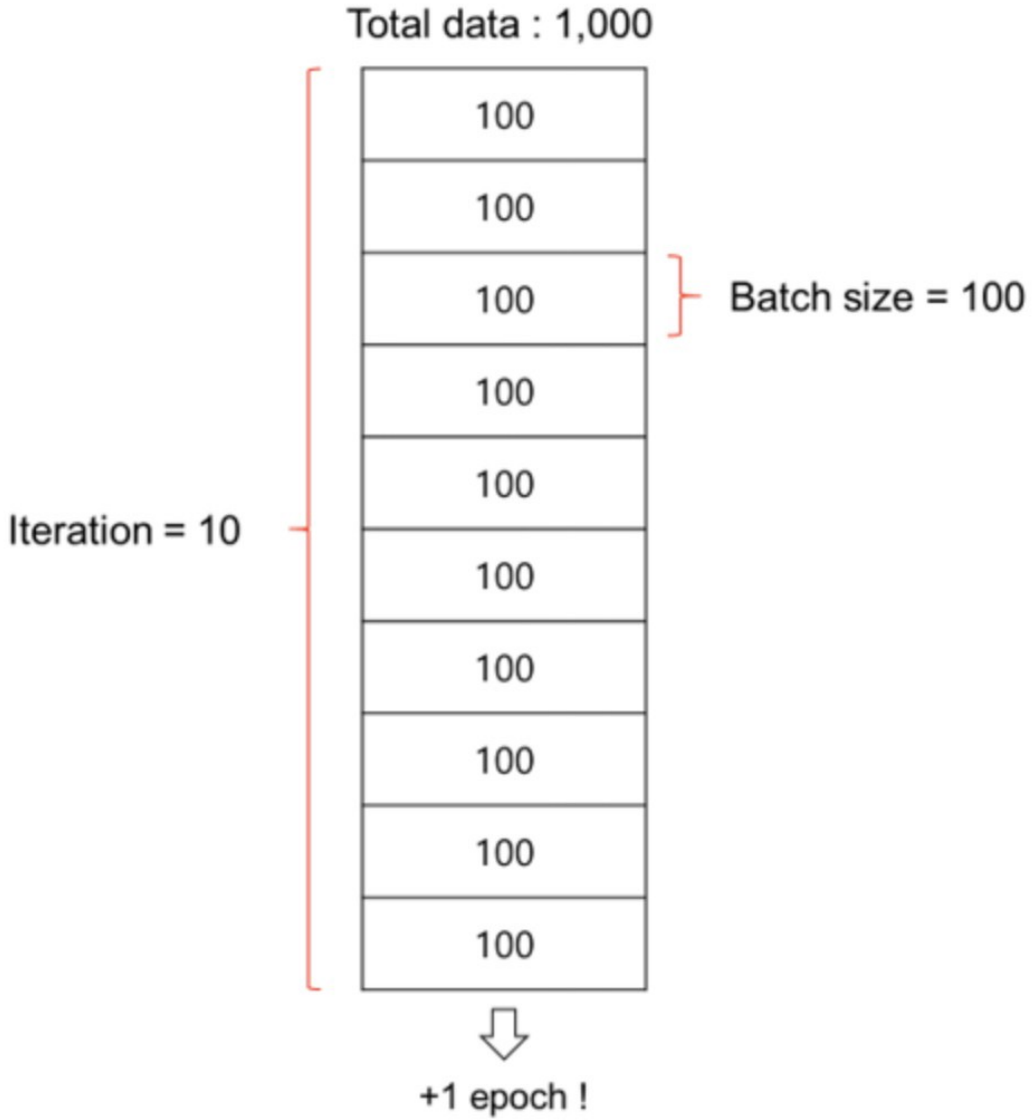
EPOCH

An **epoch** is a term used in machine learning and indicates the number of passes of the entire training dataset the machine learning algorithm has completed. Datasets are usually grouped into batches (especially when the amount of data is very large).



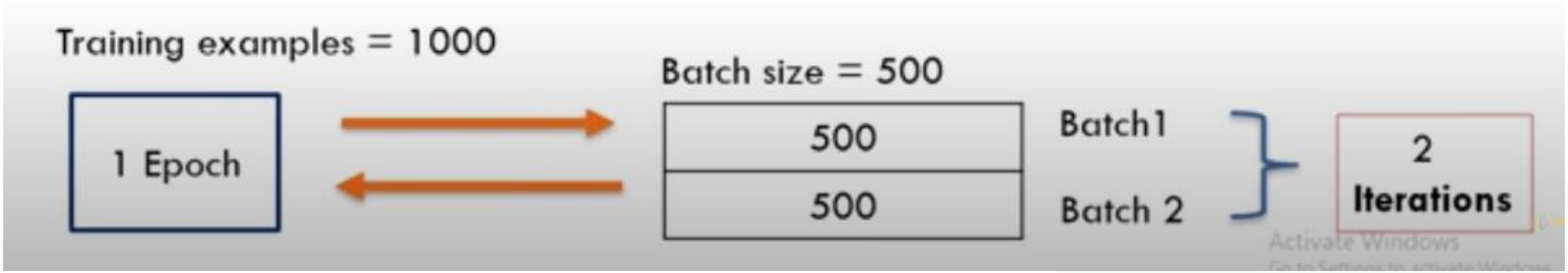
BATCHSIZE

Batch size is a term used in machine learning and refers to **the number of training examples utilized in one iteration**.



DIFFERENCE BETWEEN BATCH AND EPOCH

Example: if you have 1000 training examples, and your batch size is 500, then it will take 2 iterations to complete 1 epoch.



EPOCH

EPOCH

```
model.fit(x=X_train,y=y_train.values,
          validation_data=(X_test,y_test.values),
          batch_size=128,epochs=400)
```

Epoch

One epoch means, the entire dataset is passed forward and backward through the neural network once.

EPOCH

```
model.fit(x = X_train, y = y_train, batch_size = 32, epochs = 300)
```

```
Epoch 1/300
22/22 [==============================] - 4s 2ms/step - loss: 256629.3281
Epoch 2/300
22/22 [==============================] - 0s 2ms/step - loss: 256489.2812
Epoch 3/300
22/22 [==============================] - 0s 3ms/step - loss: 256325.5469
```

700 (TRAIN DATASI) / 32 (BATCH SIZE) = 22

700 ÷ 32 =

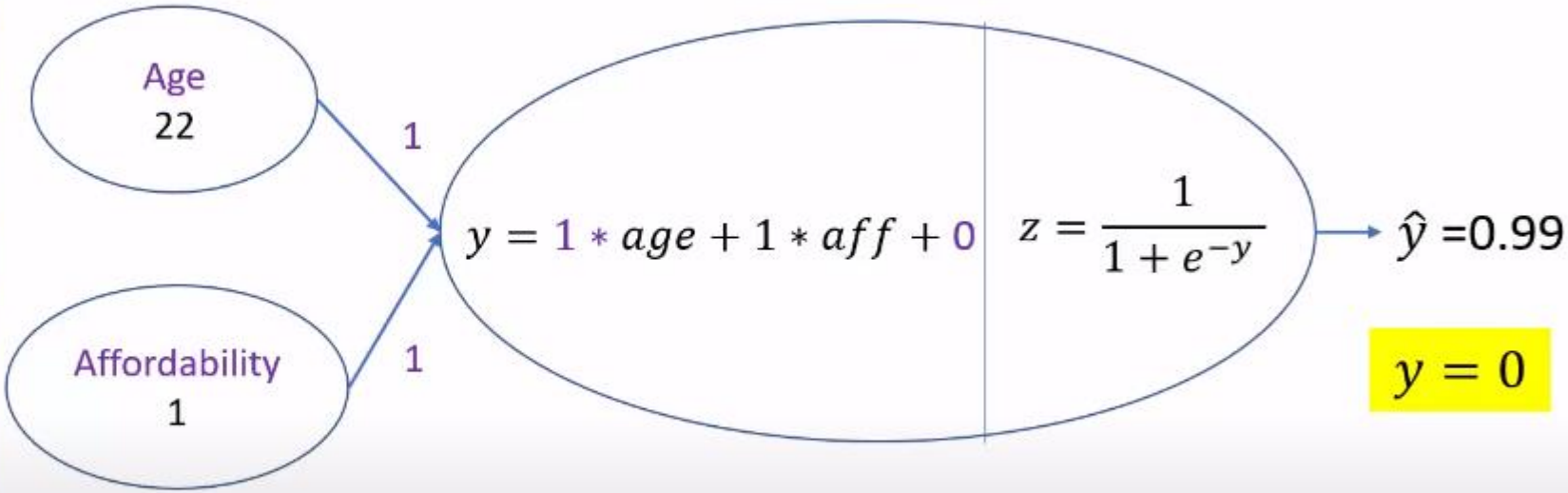
21,875

EPOCH

age	affordability	have_insurance
22	1	0
25	0	0
47	1	1
52	0	0
46	1	1
56	1	1
55	0	0
60	0	1
62	1	1
61	1	1
18	1	0
28	1	0
27	0	1

$$y = w1 * x1 + w2 * x2 + bias$$

Age Affordability

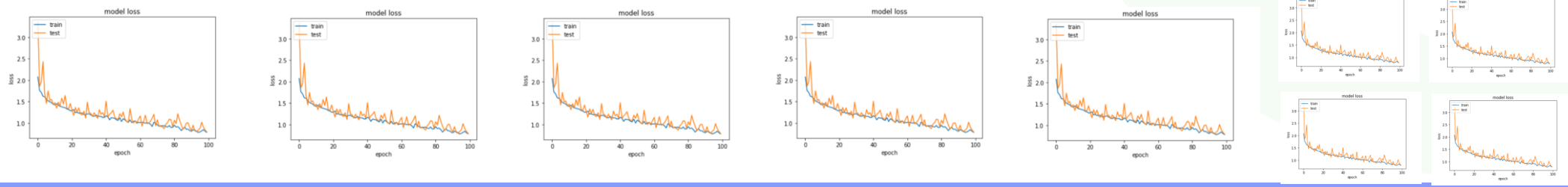


$$error1 = -(y\log(\hat{y}) + (1-y)\log(1-\hat{y}))$$
$$= 4.6$$

700 (TRAIN DATASI) / 32 (BATCH SIZE) = 22

1 EPOCH

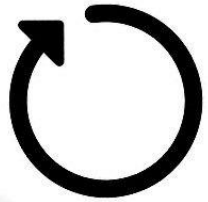
22 iteration



DIFFERENCE BETWEEN BATCH AND EPOCH

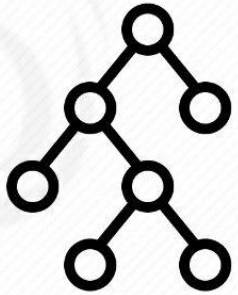
Epoch :

An Epoch represent one iteration over the entire dataset.



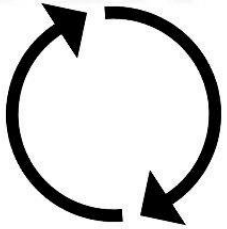
Batch :

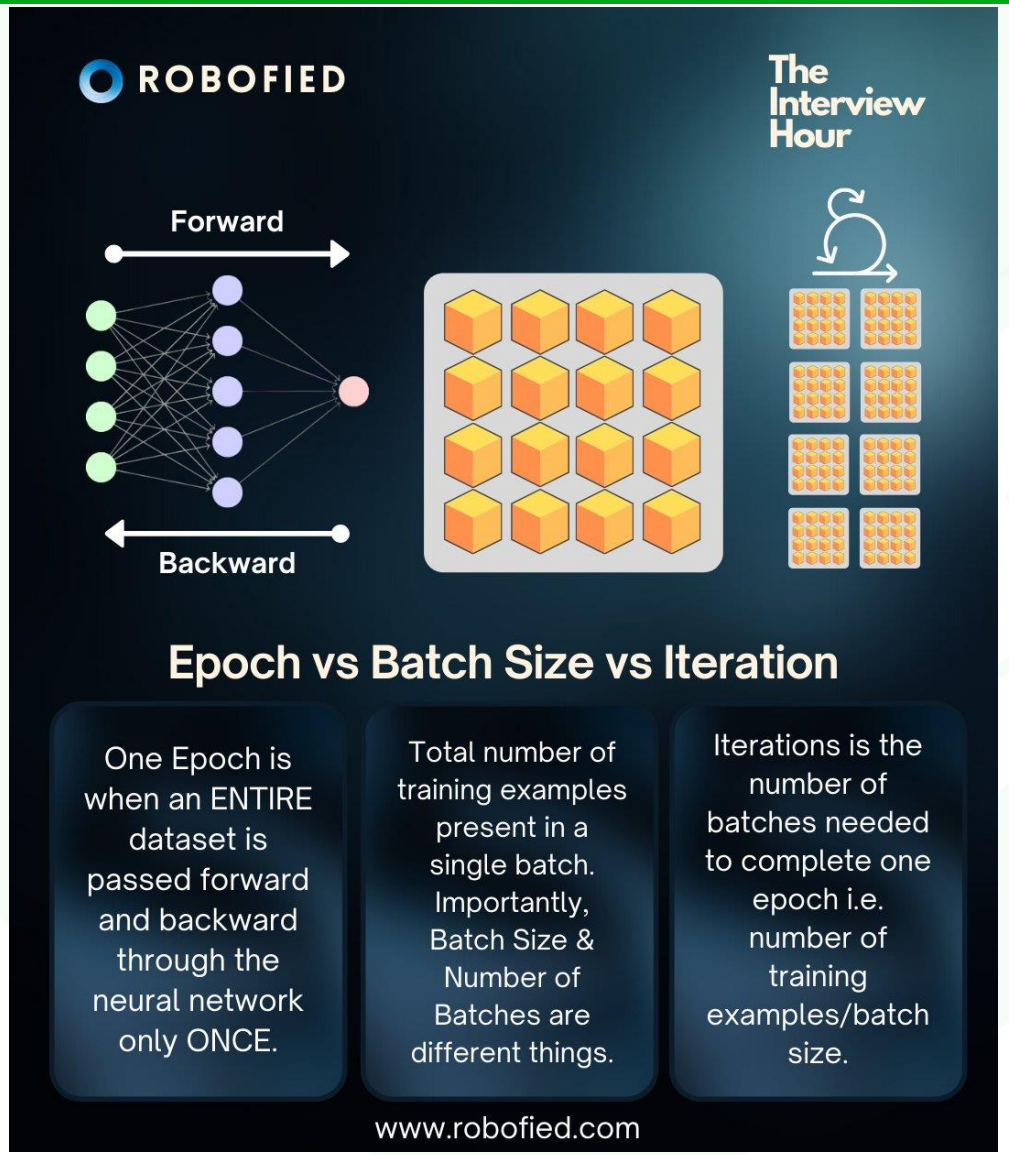
We cannot pass the entire dataset into the Neural Network at once. So, we divide the dataset into number of batches.



Iteration :

If we have 1000 images as Data ane a batch size of 20, then an Epoch should run 1000/20 = 50 iteration.





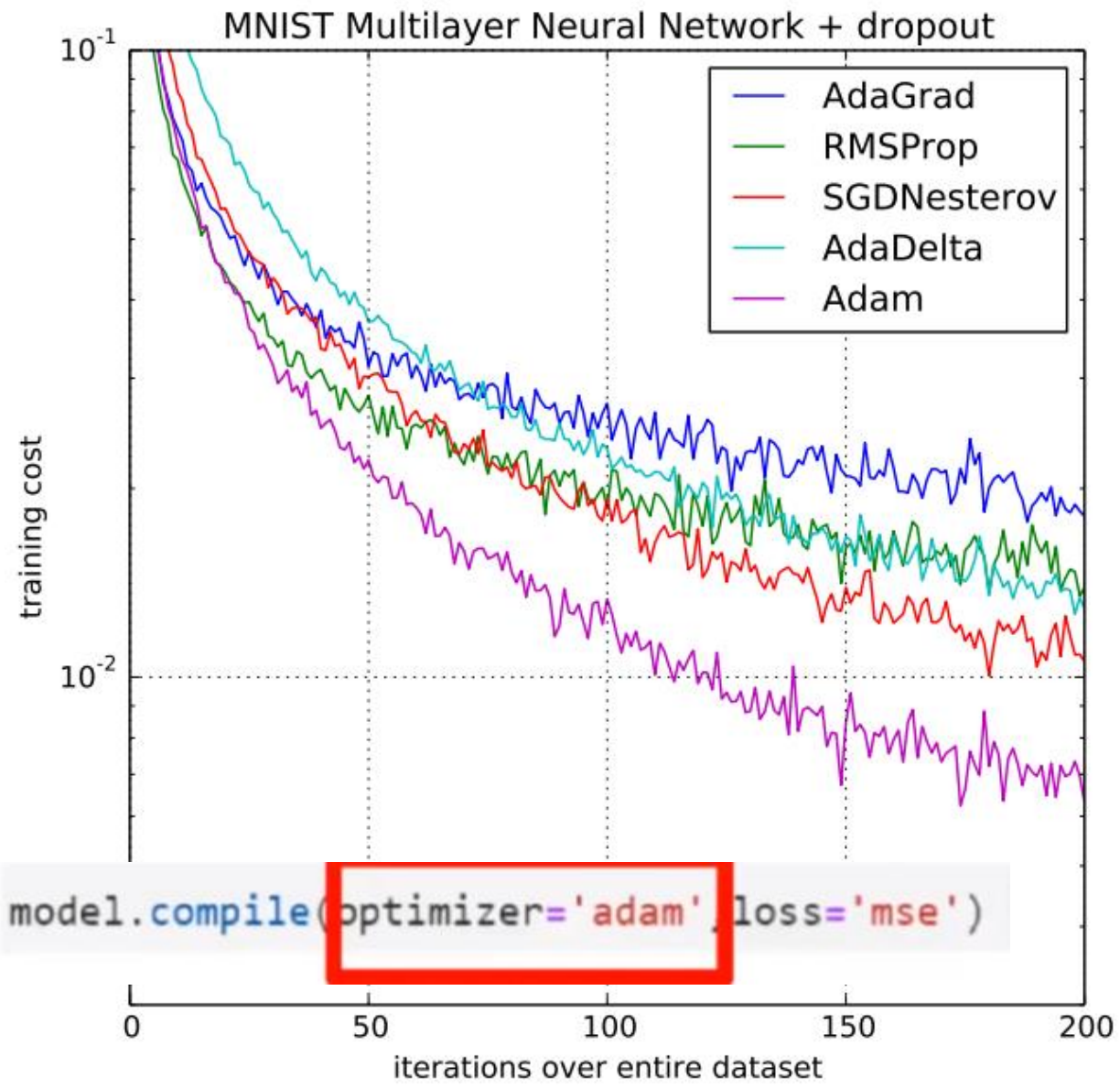
GRADIENT DESCENT

GRADIENT DESCENT

GRADIENT DESCENT

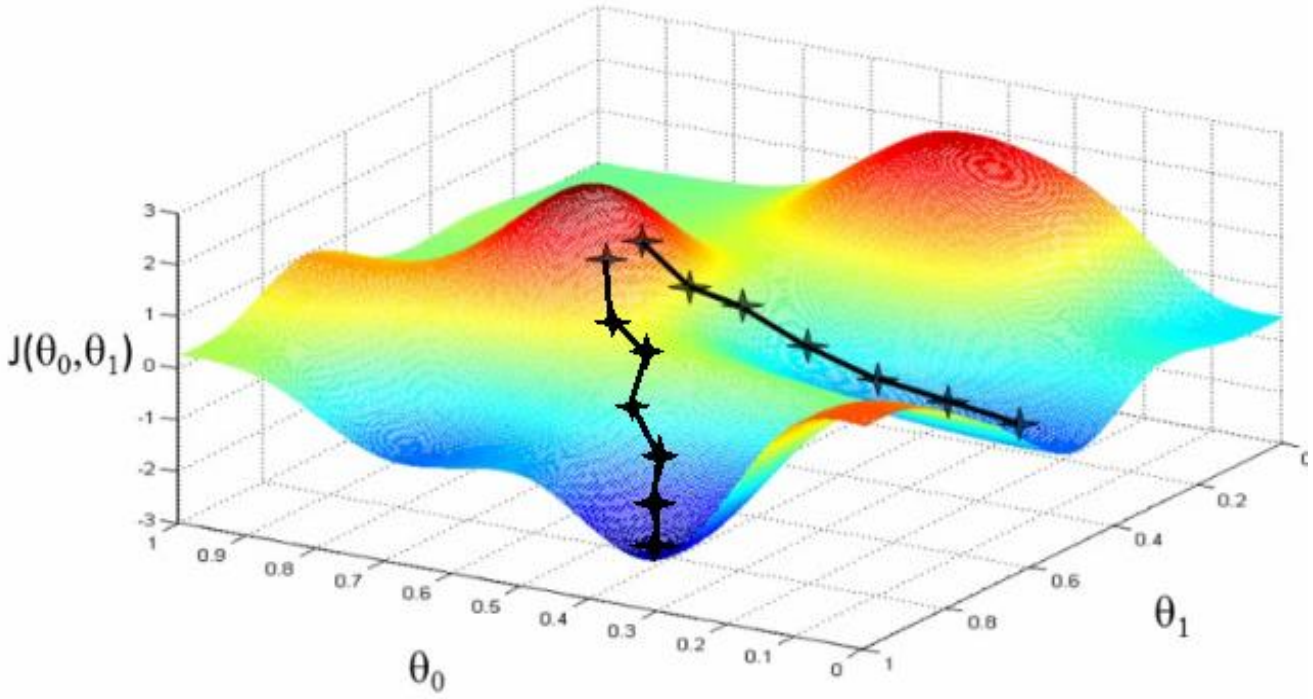
OPTIMIZER

Gradient descent is an optimization algorithm that uses the gradient of the objective function to navigate the search space. Optimization is a mathematical discipline that determines the “best” solution in a quantitatively well-defined sense.



GRADIENT DESCENT

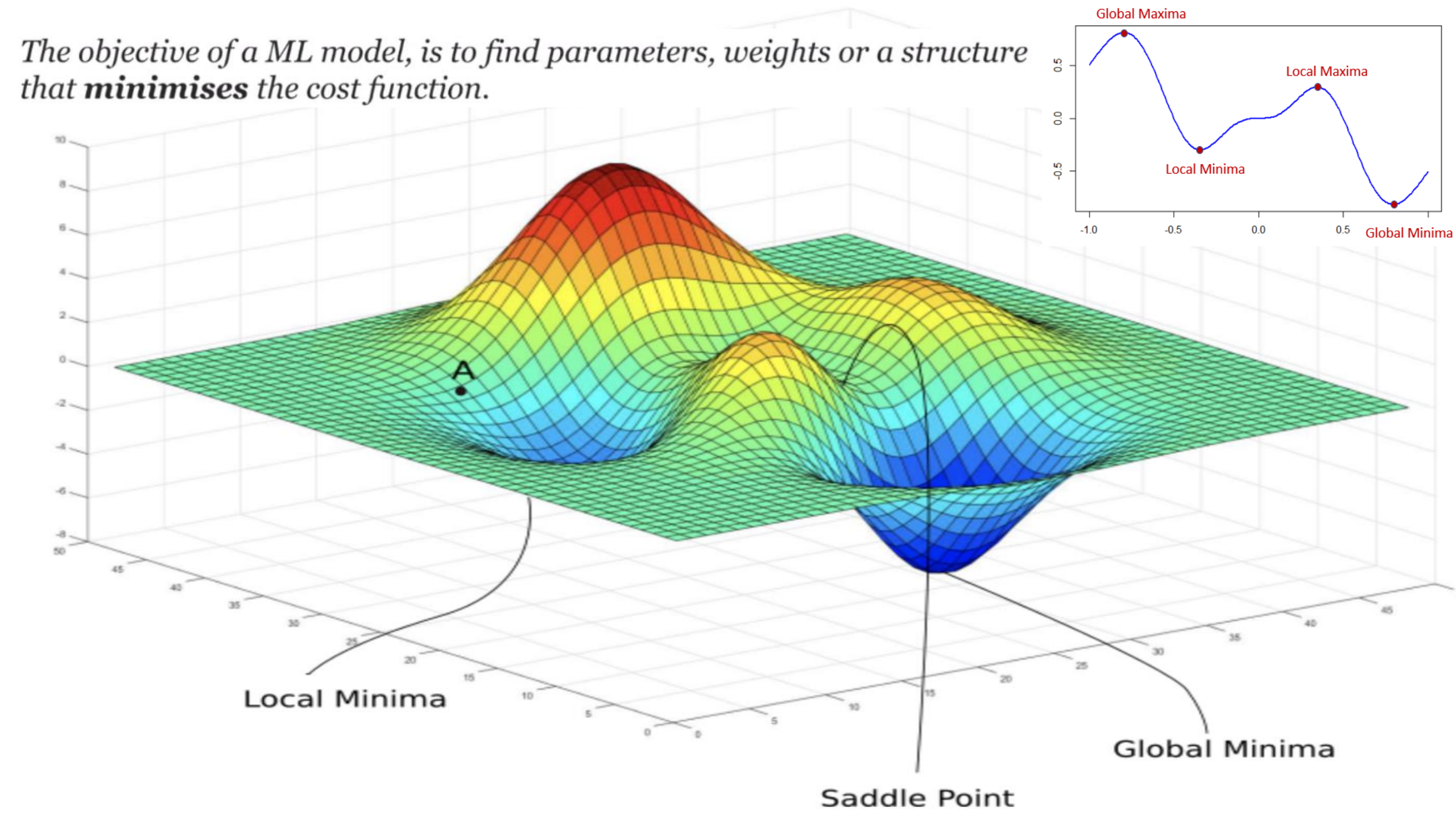
Gradient descent is the process of using gradients to find the minimum value of the cost function, while backpropagation is calculating those gradients by moving in a backward direction in the neural network.



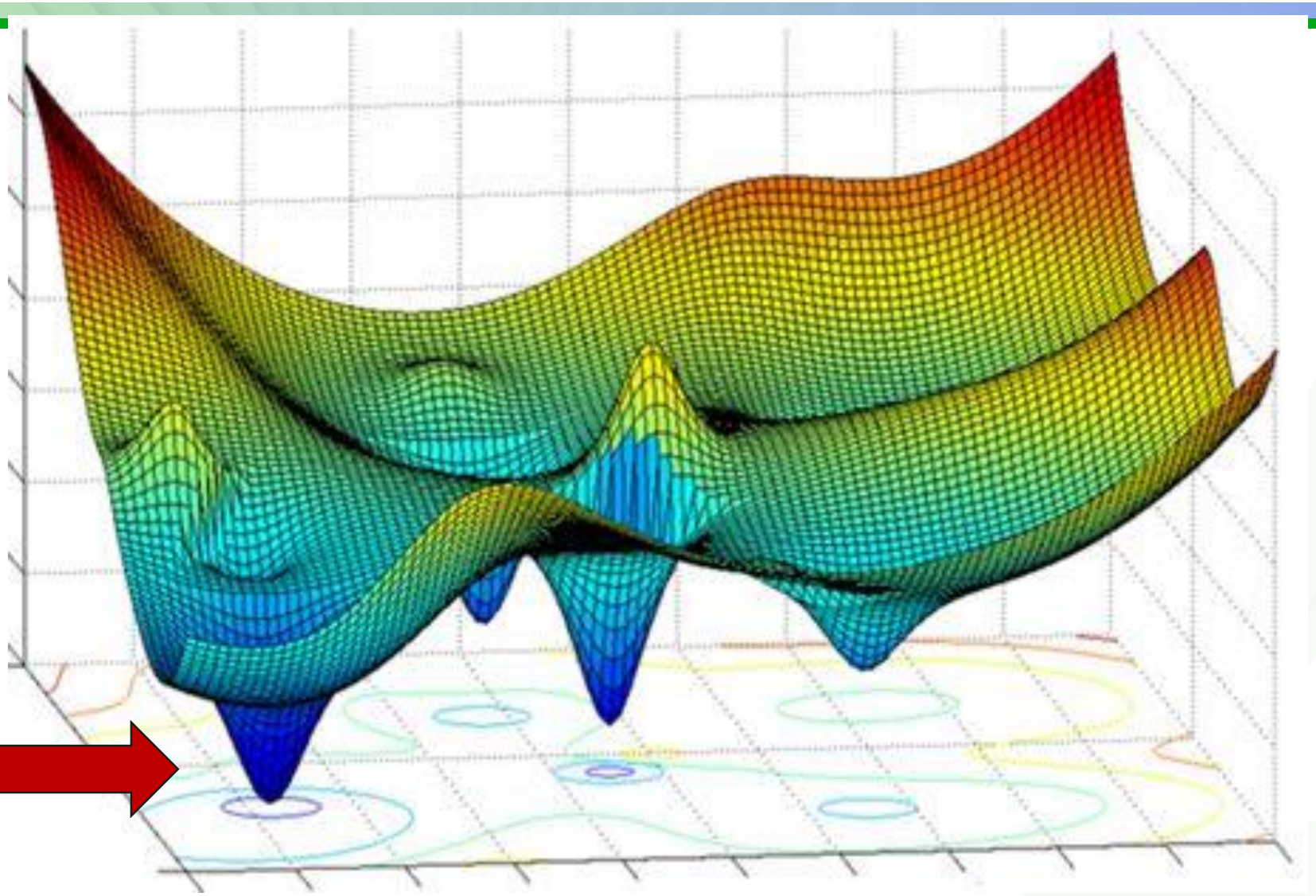
GRADIENT DESCENT

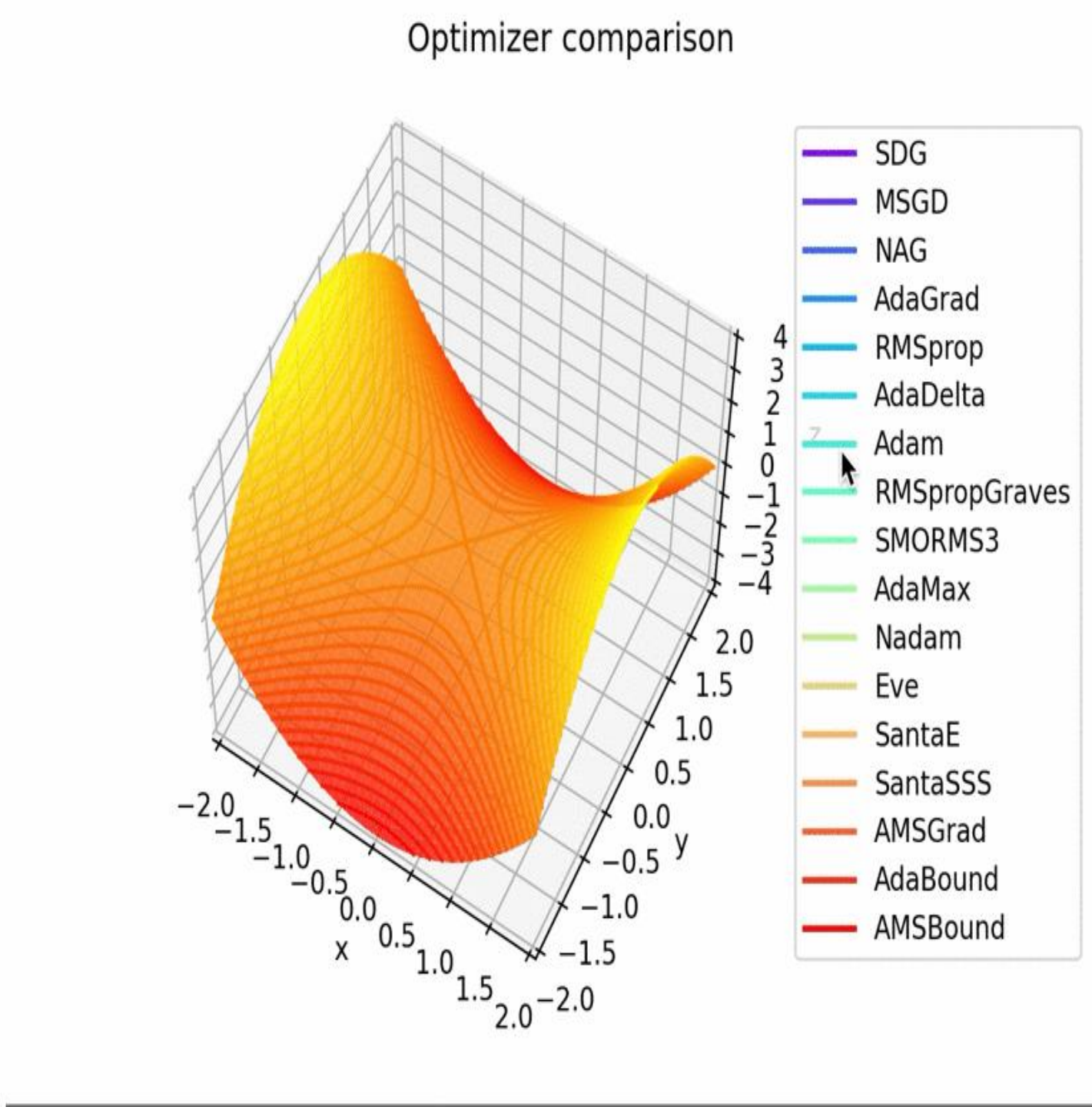

Leo Learns the Ropes
at Project Chimps

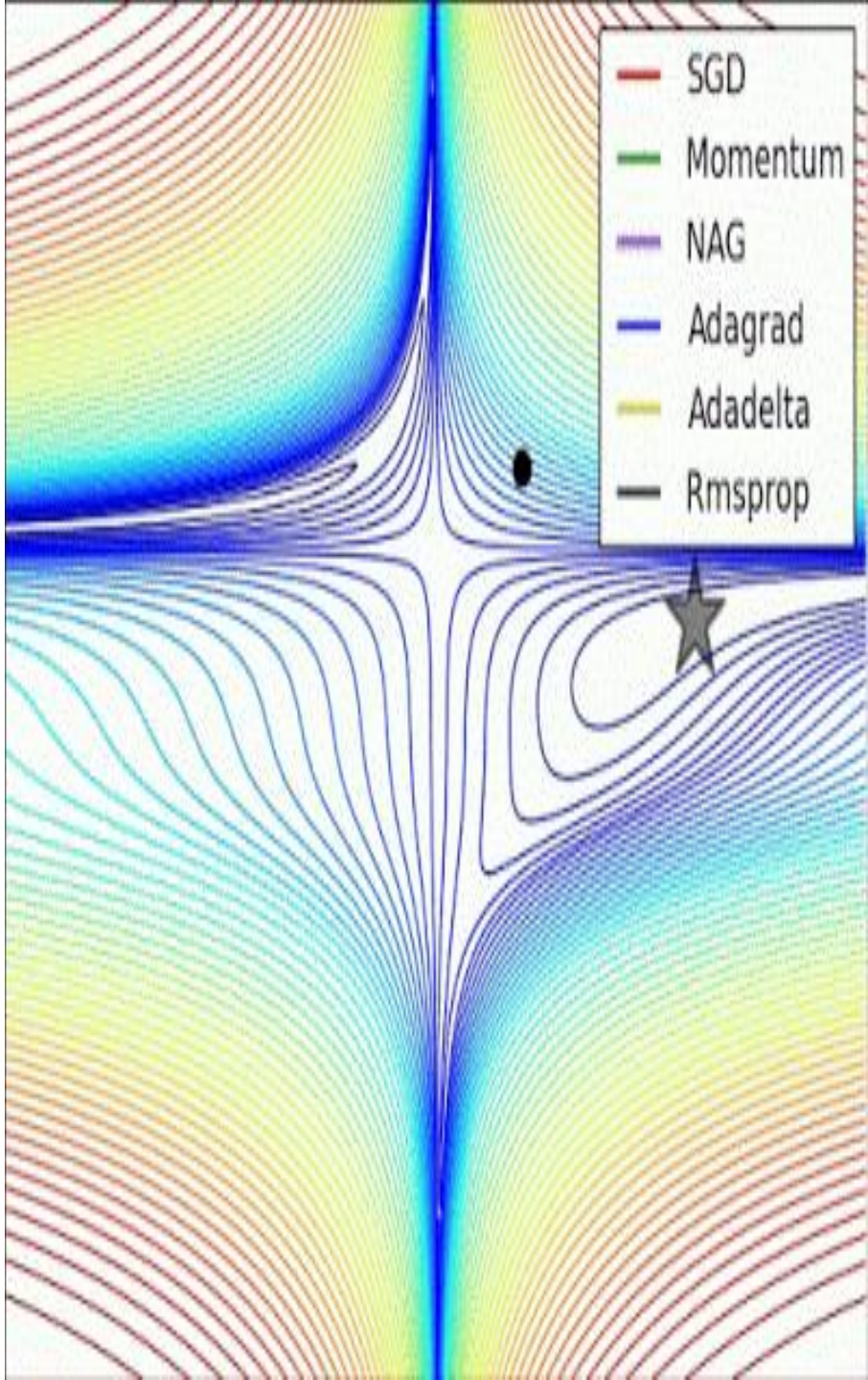
*The objective of a ML model, is to find parameters, weights or a structure that **minimises** the cost function.*



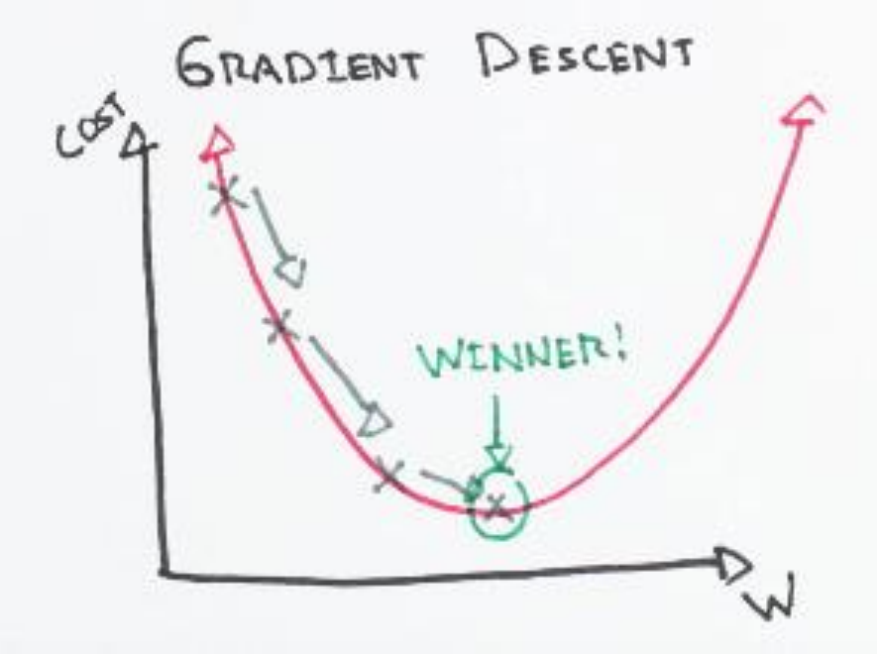
GRADIENT DESCENT



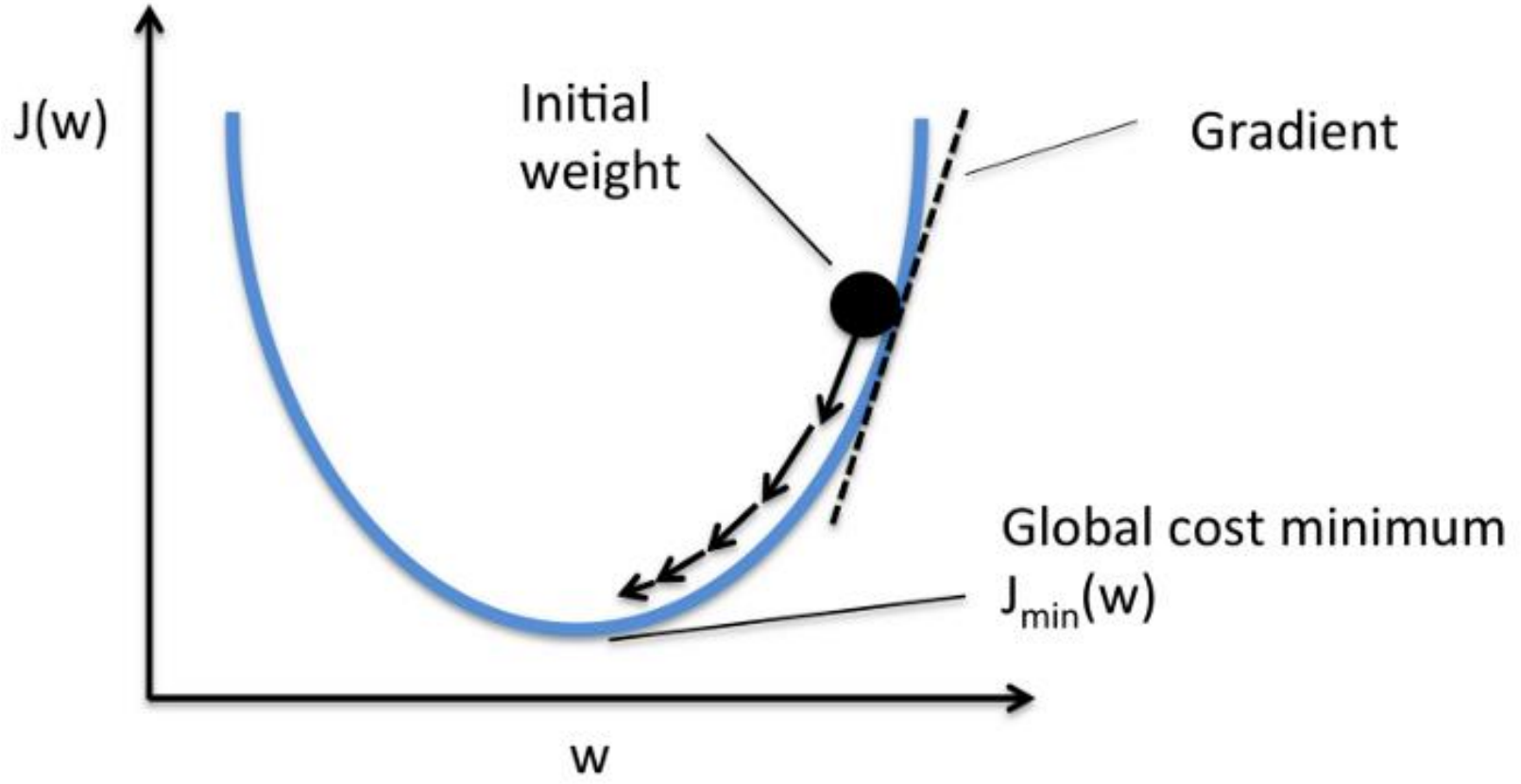

Optimizer comparison
SDG
MSGD
NAG
AdaGrad
RMSprop
AdaDelta
Adam
RMSpropGraves
SMORMS3
AdaMax
Nadam
Eve
SantaE
SantaSSS
AMSGrad
AdaBound
AMSBound
x
y
z


SGD
Momentum
NAG
Adagrad
Adadelta
Rmsprop

GRADIENT DESCENT

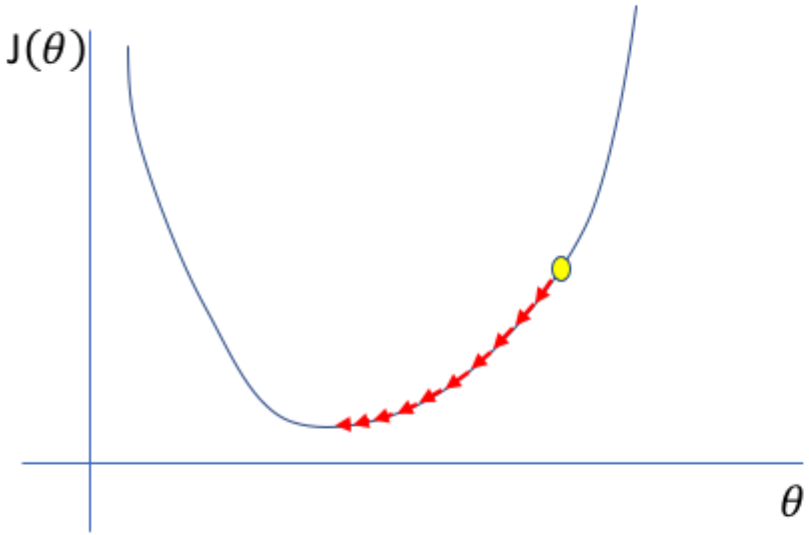


Step Size = slope x learning rate



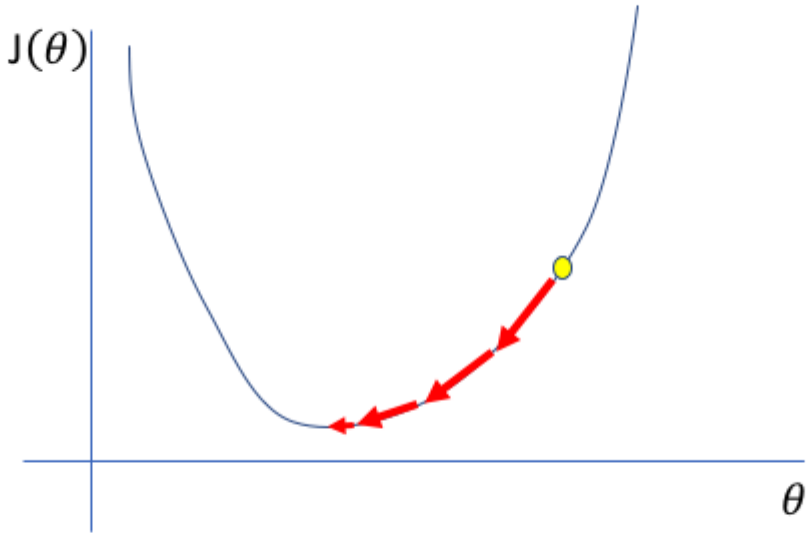
LEARNING RATE

Too low



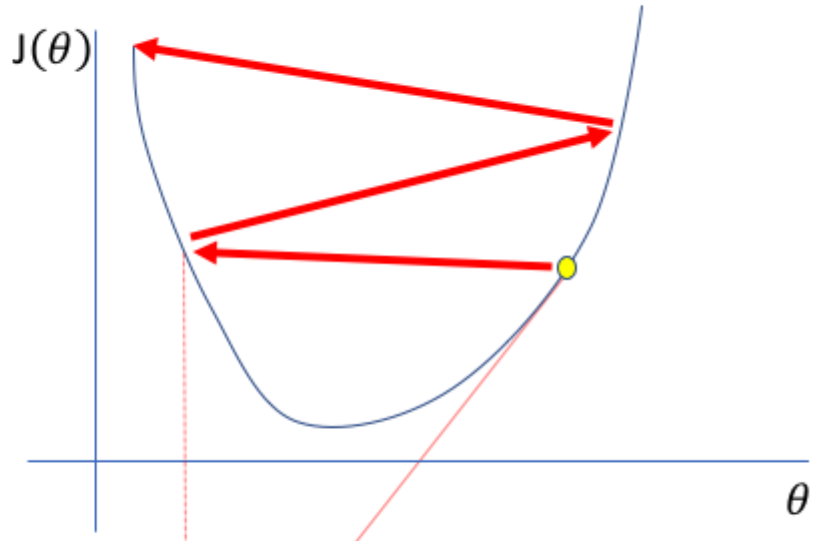
A small learning rate requires many updates before reaching the minimum point

Just right



The optimal learning rate swiftly reaches the minimum point

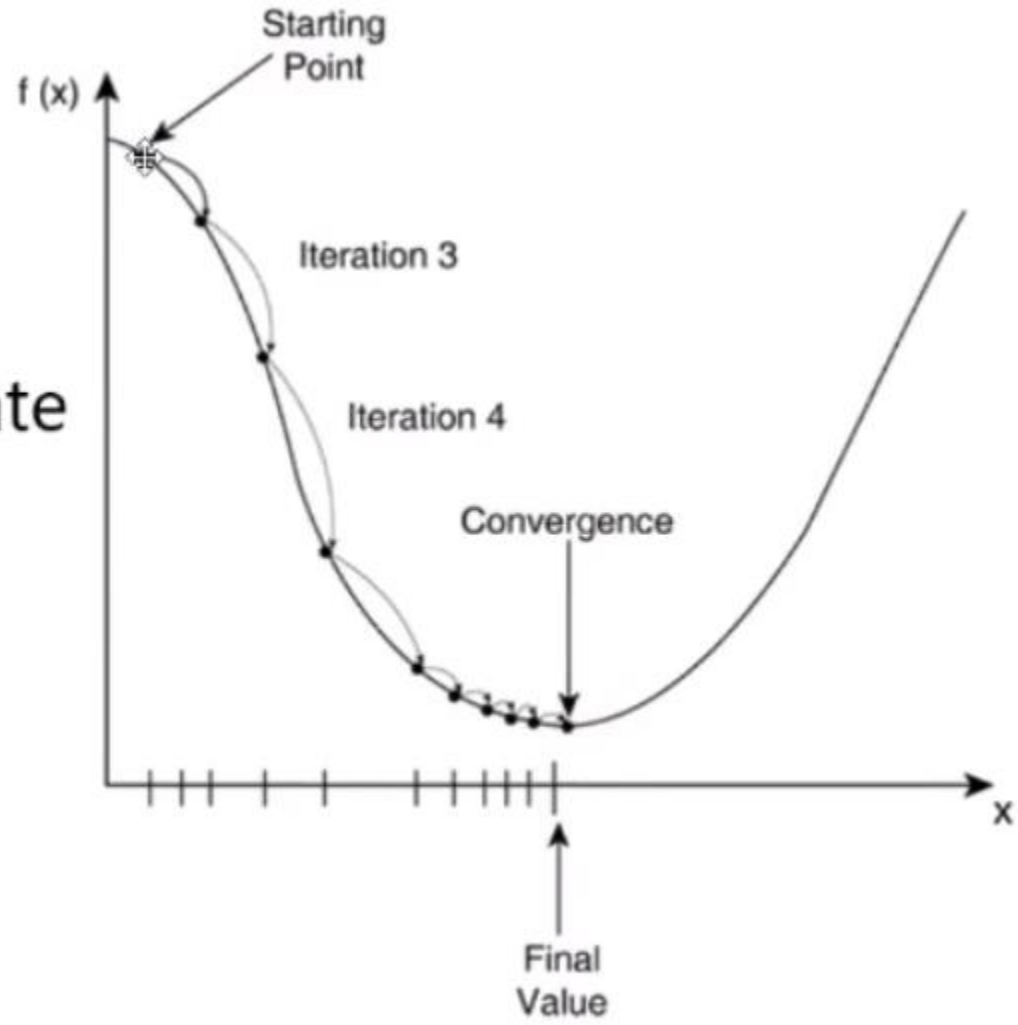
Too high



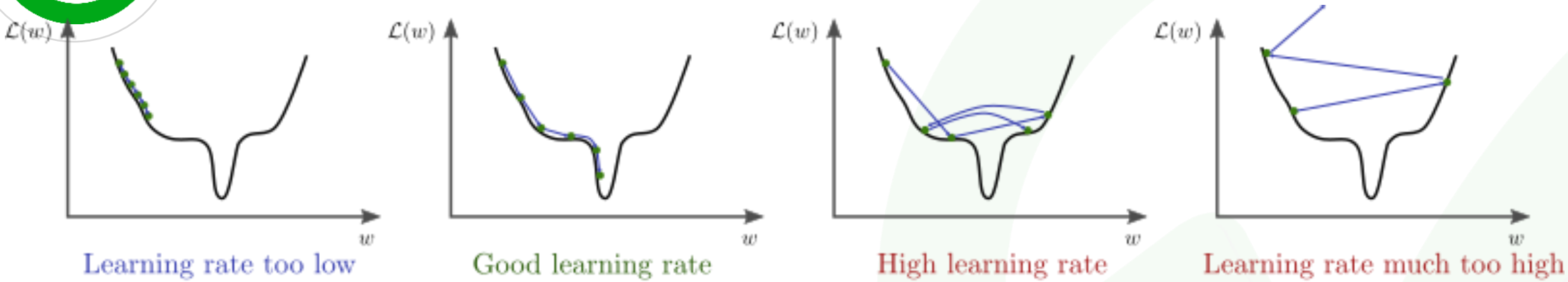
Too large of a learning rate causes drastic updates which lead to divergent behaviors

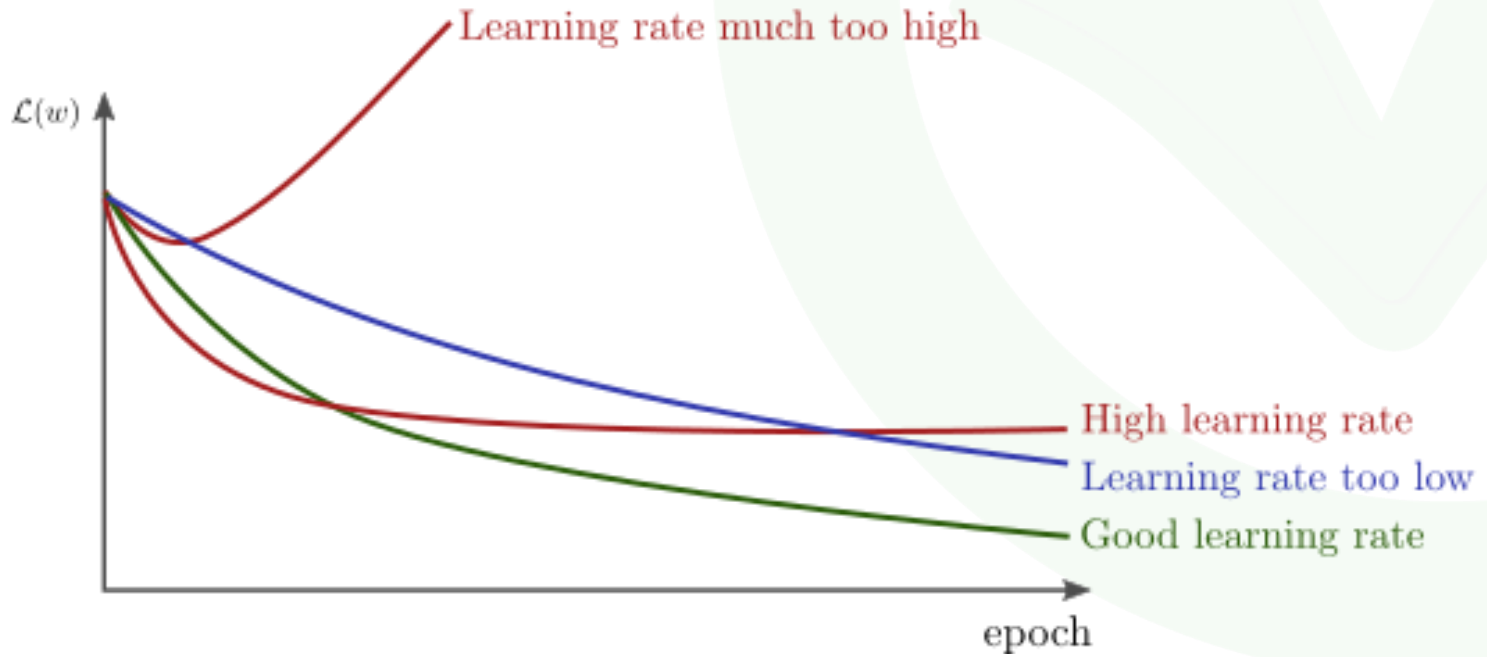
LEARNING RATE

Step Size = slope x learning rate



LEARNING RATE

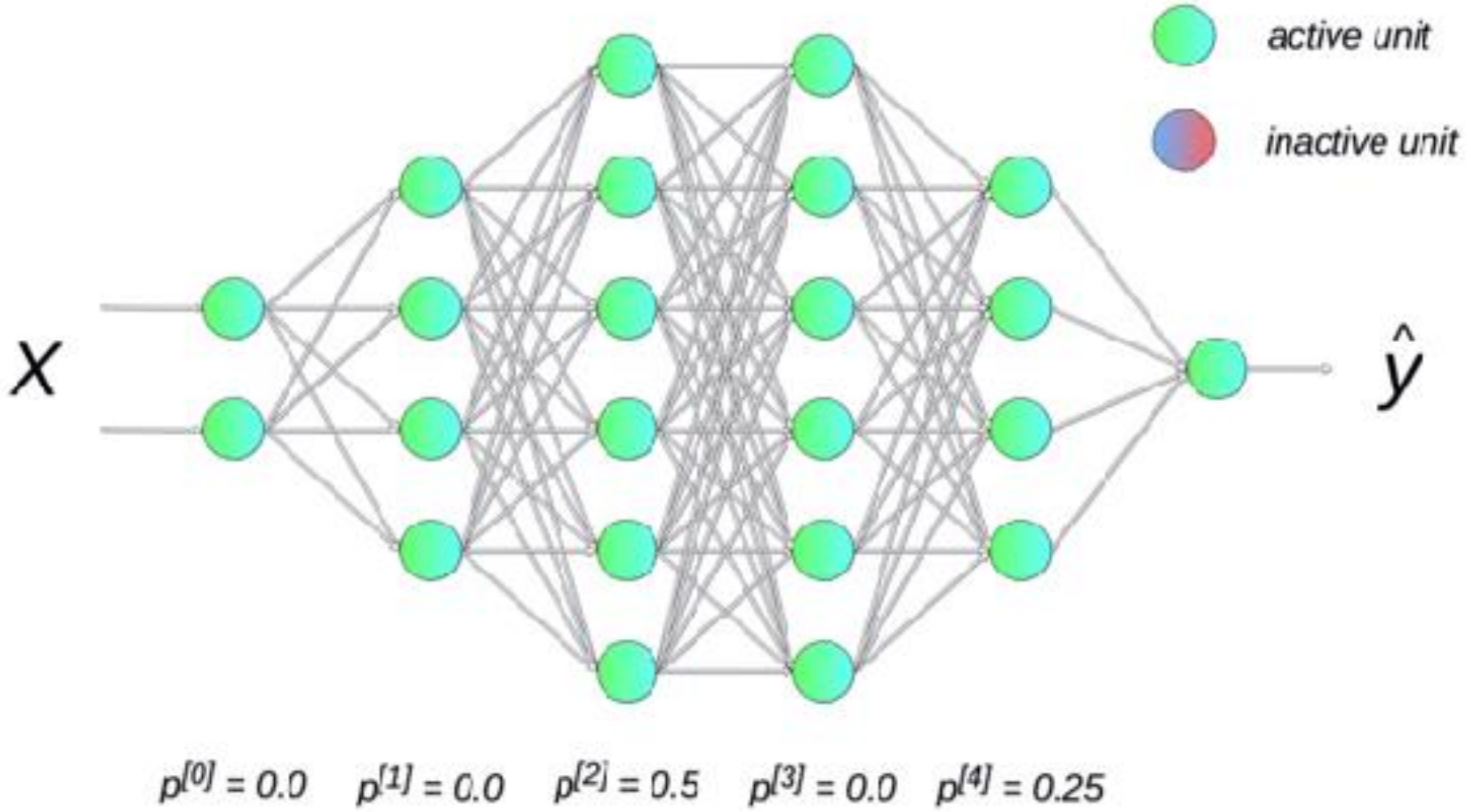




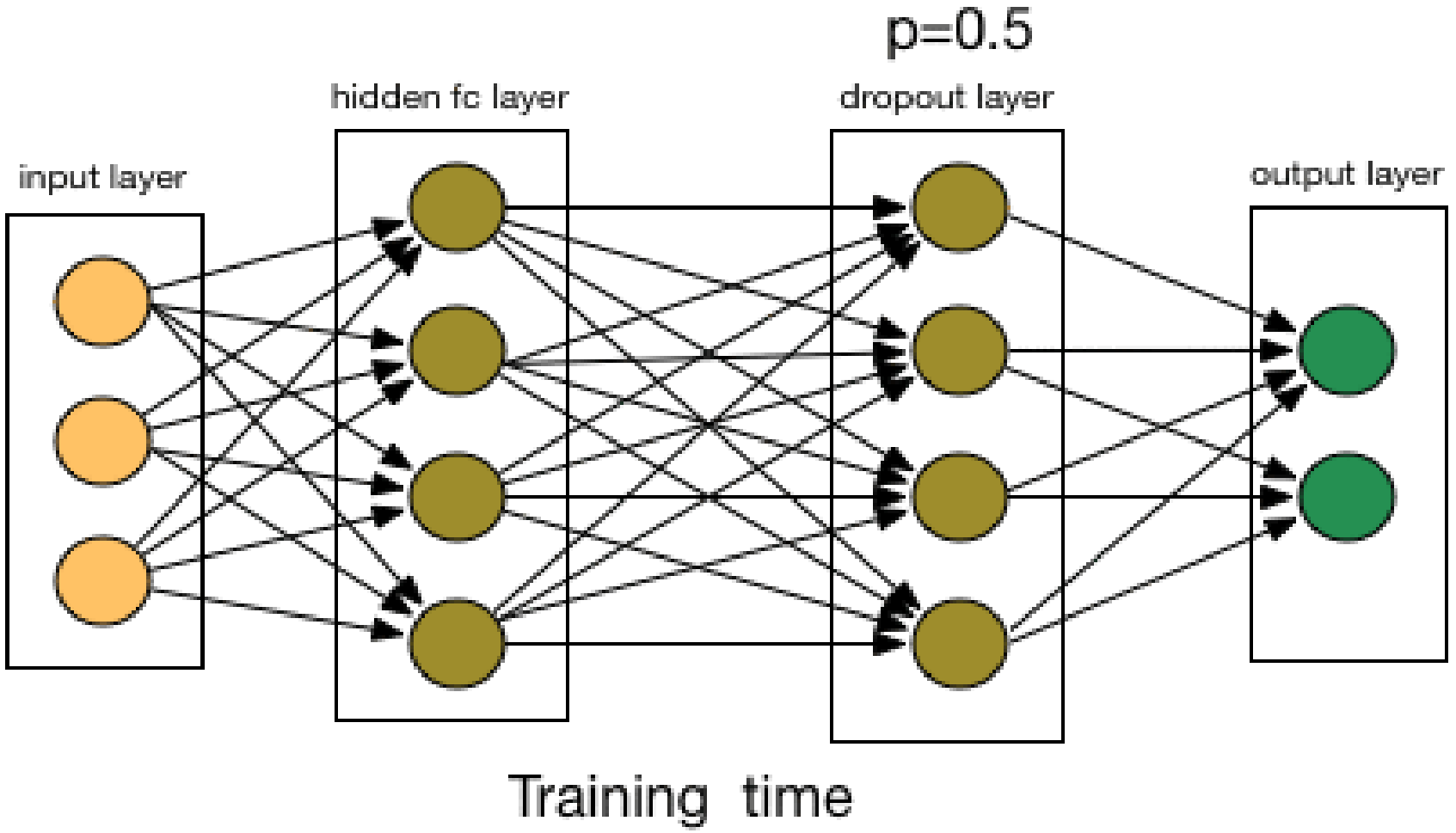
REGULARIZATION

DROPOUT
EARLYSTOPPING
L1
L2

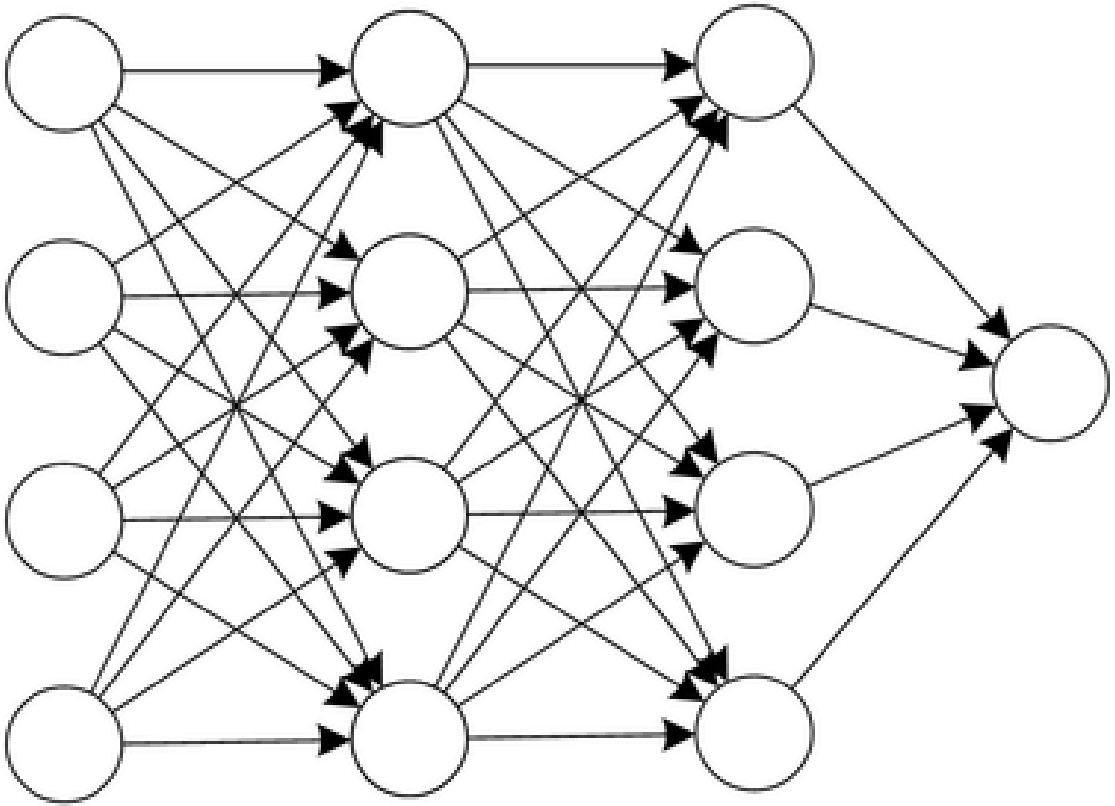
DROPOUT



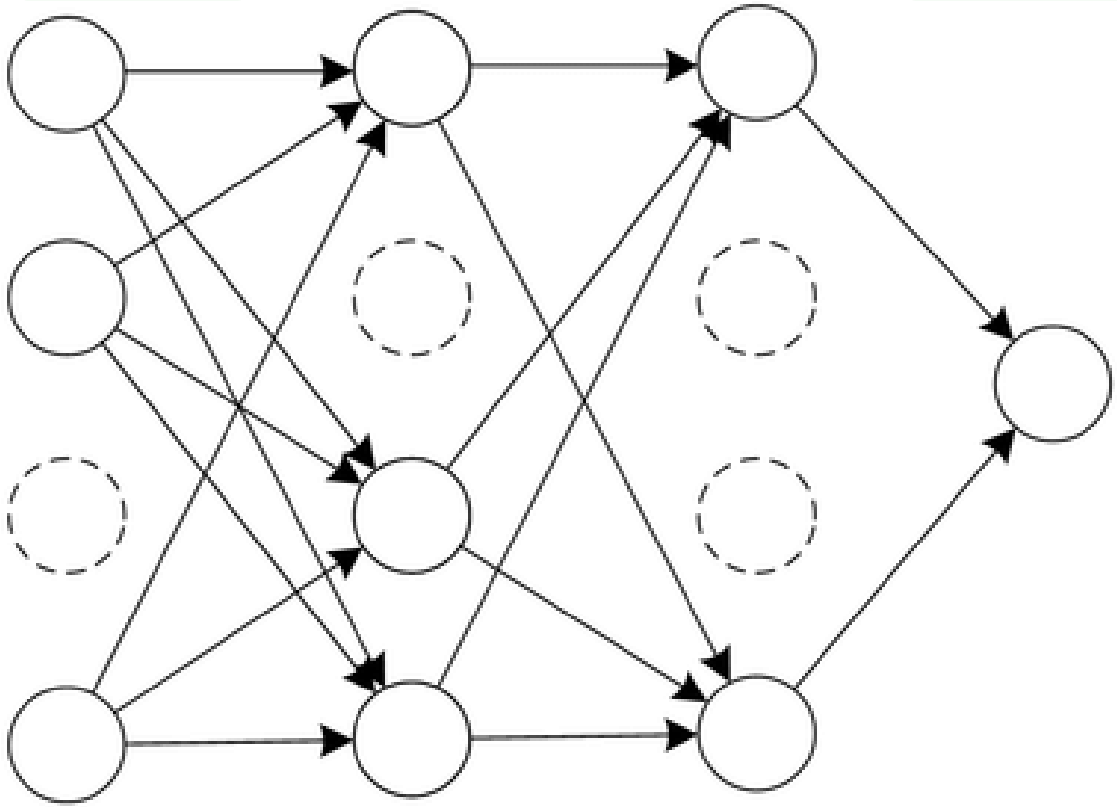
DROPOUT


p=0.5
hidden fc layer
dropout layer
input layer
output layer
Training time

DROPOUT

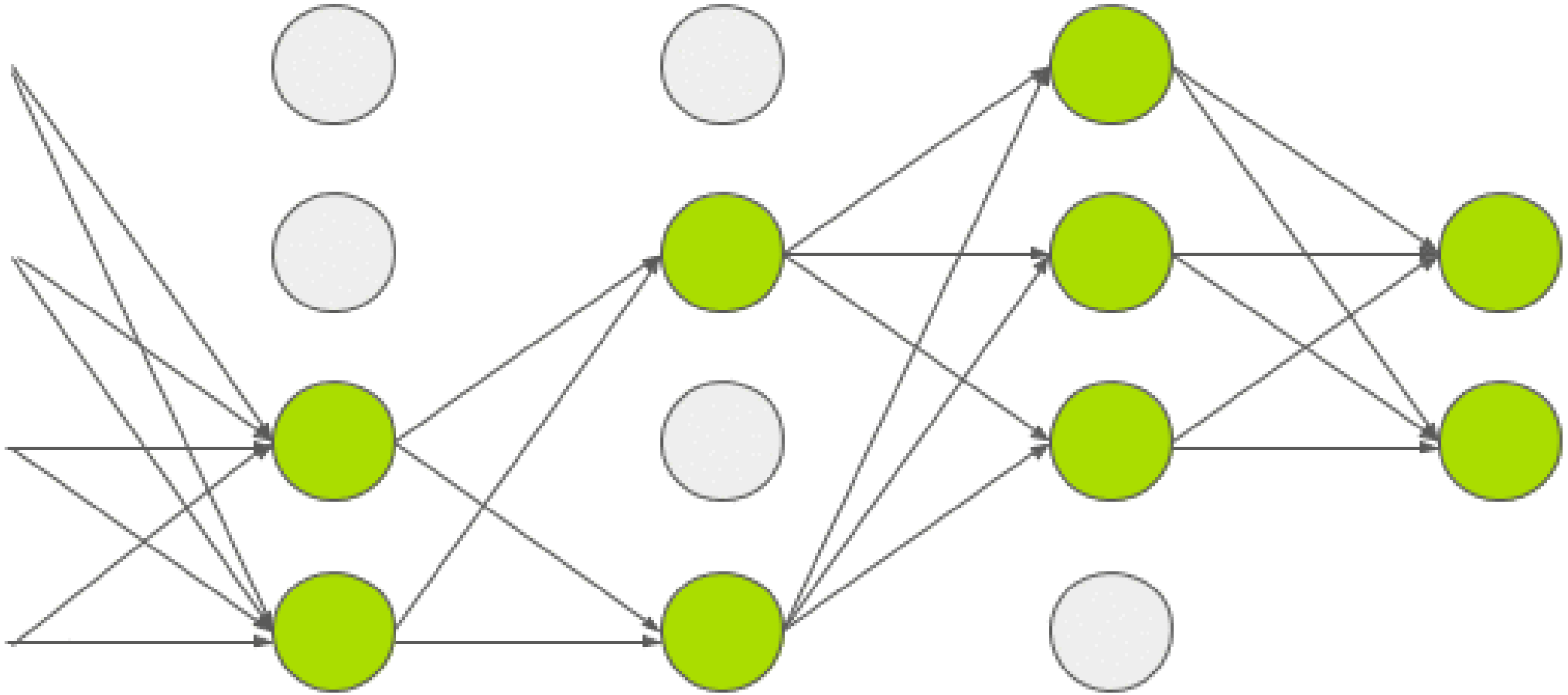


(a) Standard Neural Network



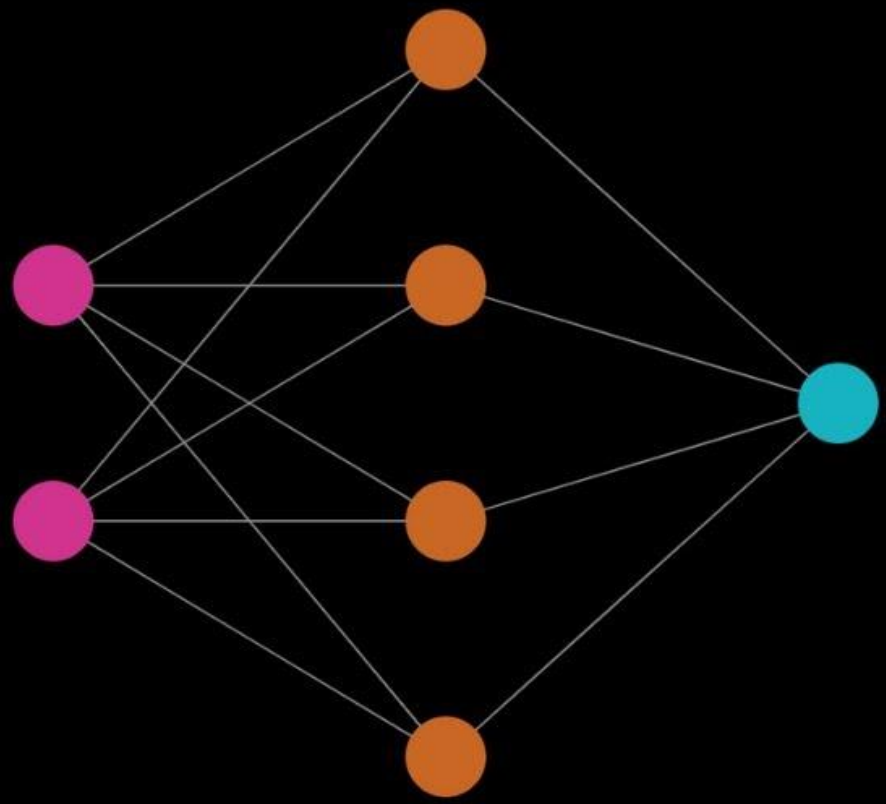
(b) Network after Dropout

DROPOUT

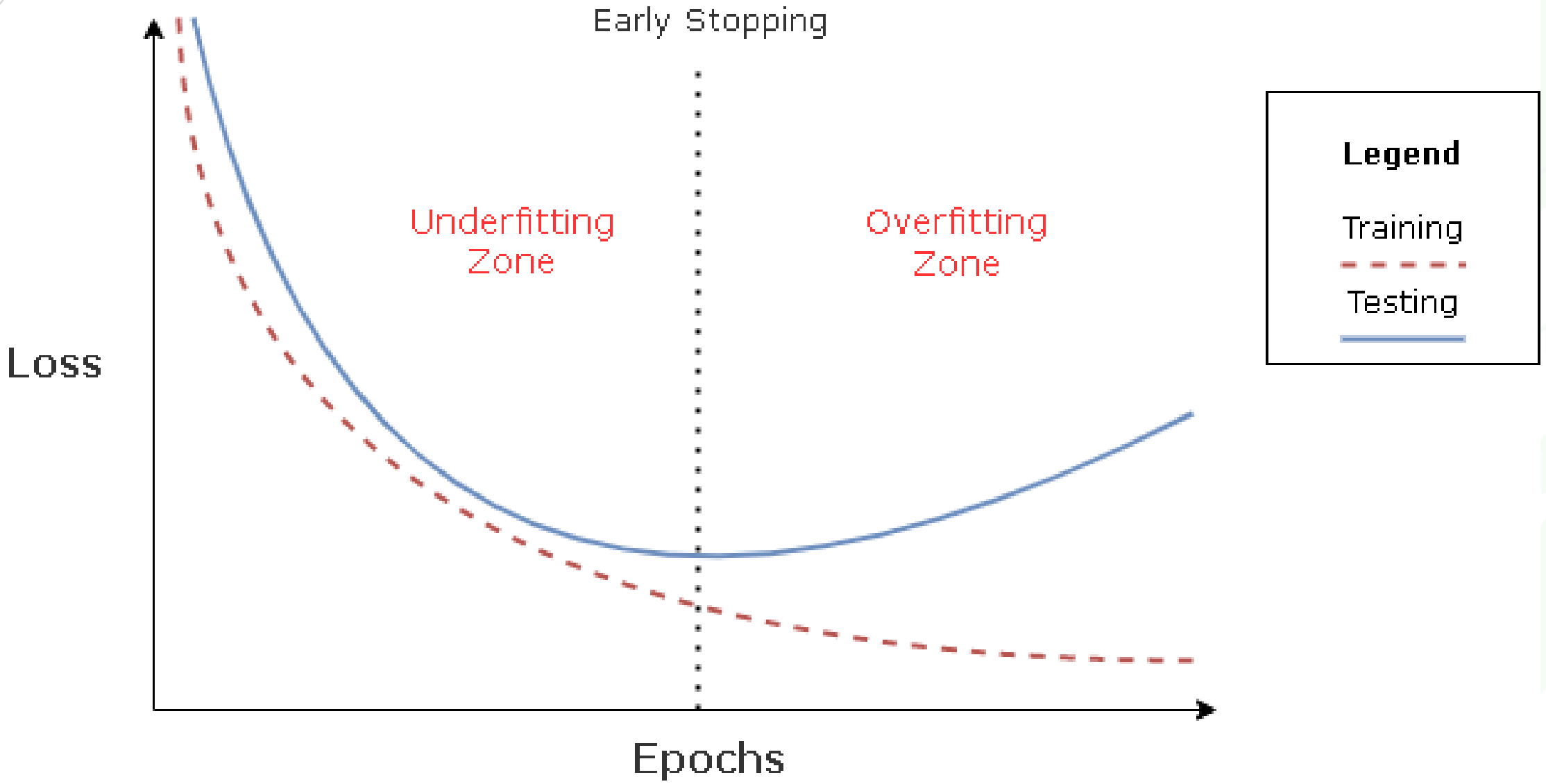


EARLYSTOPPING

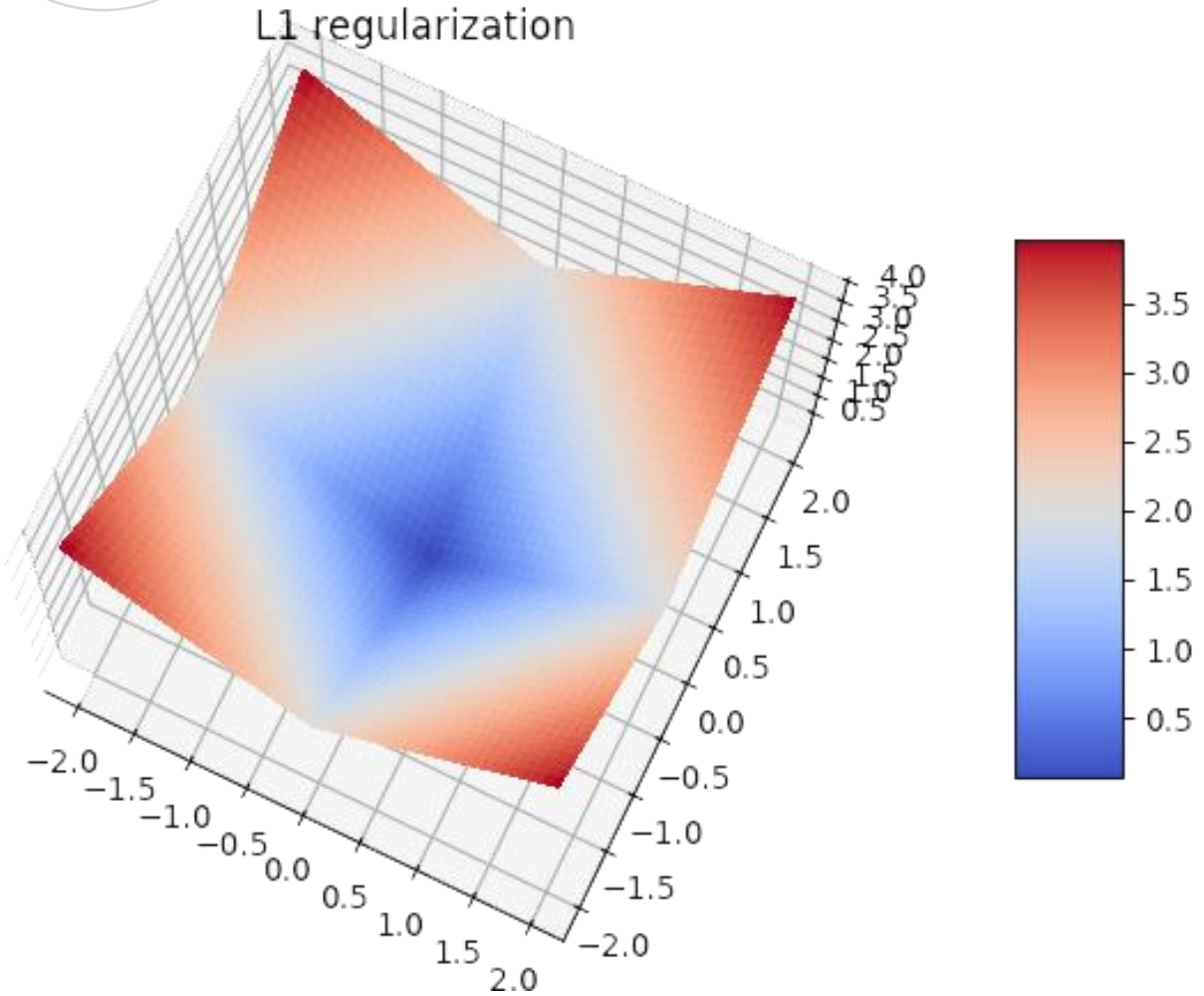
EARLY STOPPING TO PREVENT OVERFITTING

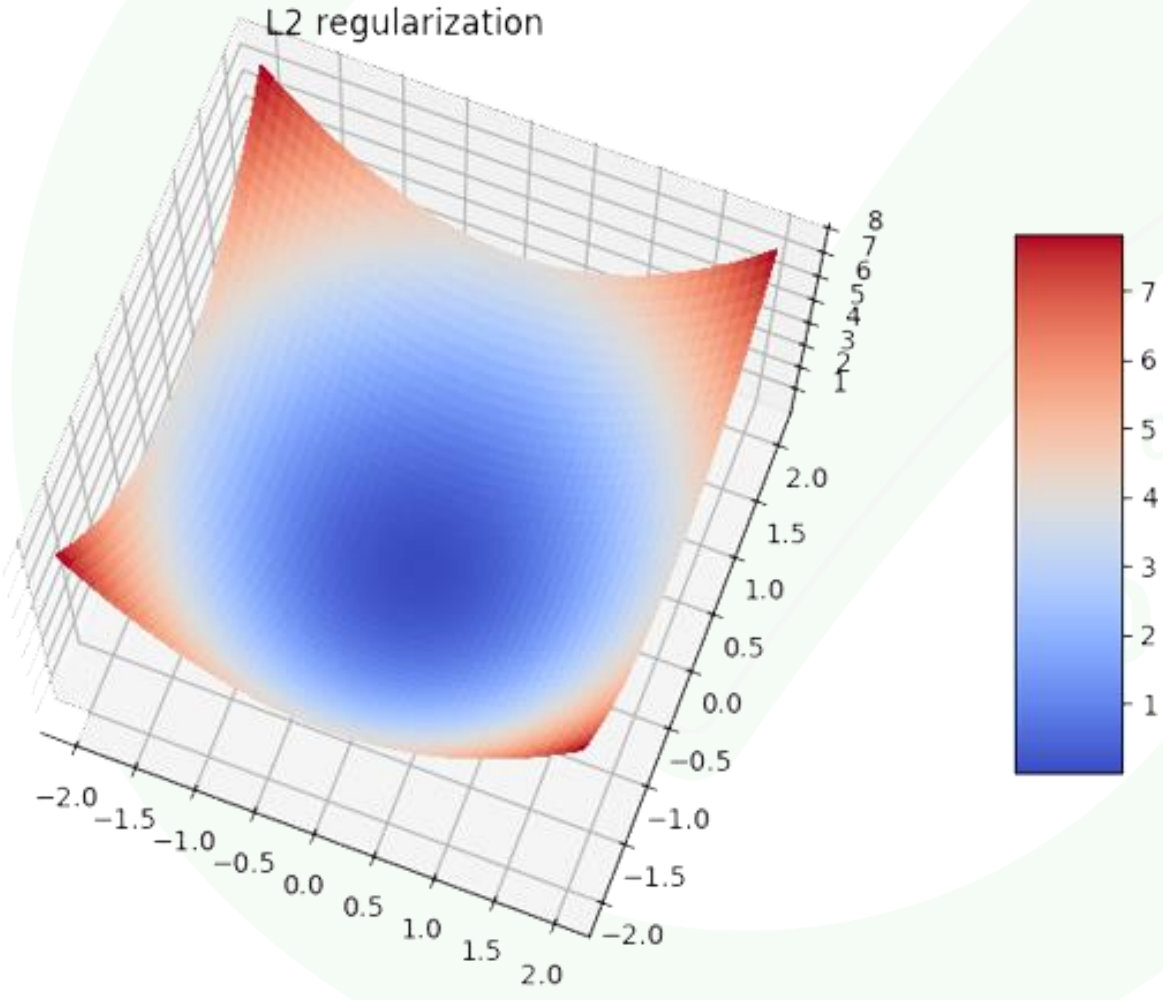


EARLYSTOPPING

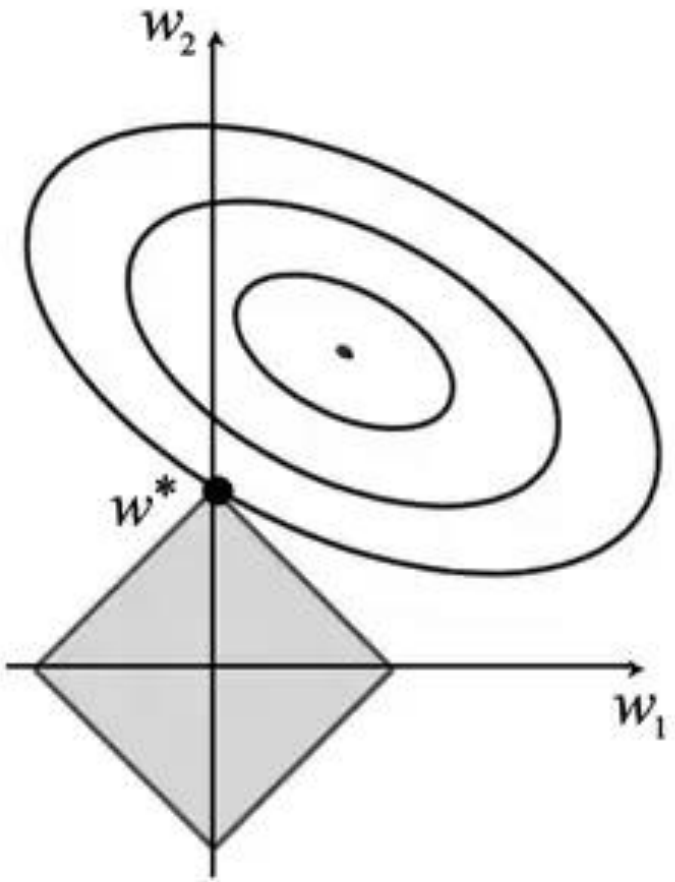


L1 and L2

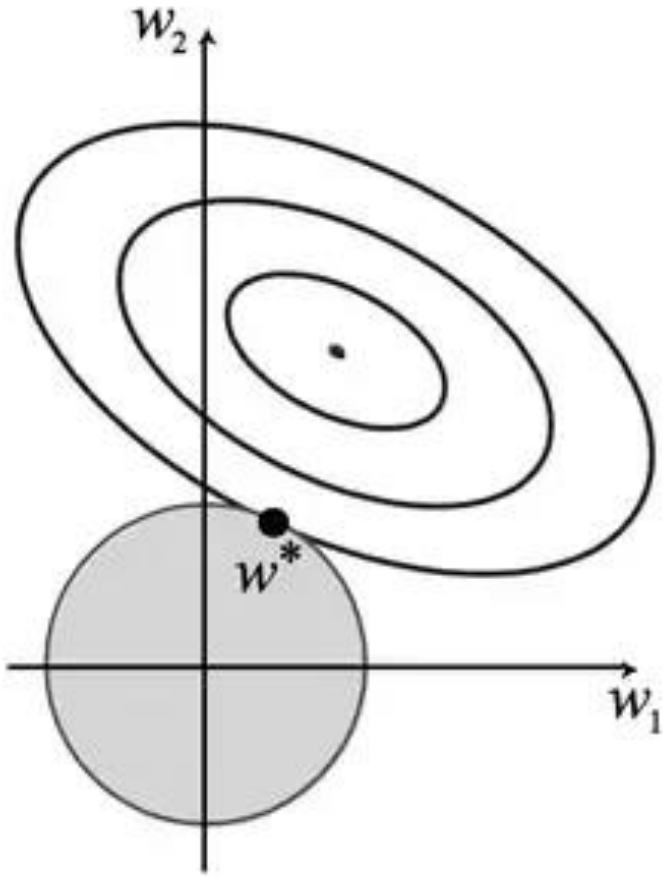

L1 regularization
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5


L2 regularization
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
8
7
6
5
4
3
2
1
7
6
5
4
3
2
1

L1 and L2



L1



L2