Training a Neural Network

Hyperparameters

Neural Networks Hyperparameters

Calculating layer outputs:

$$A = \alpha\left(W^T X + b\right)$$

activation function

# of iterations/epochs

batch size

Dataset

regularization

# of neurons in the input layer

w

b

weight initialization technique

# of neurons in the hidden layers

# of hidden layers

# of neurons in the output layer

predictions − actual value = loss/error

loss function

loss

learning rate

W_x

optimization algorithm

So you want to be **a data scientist?**

List of hyperparameters

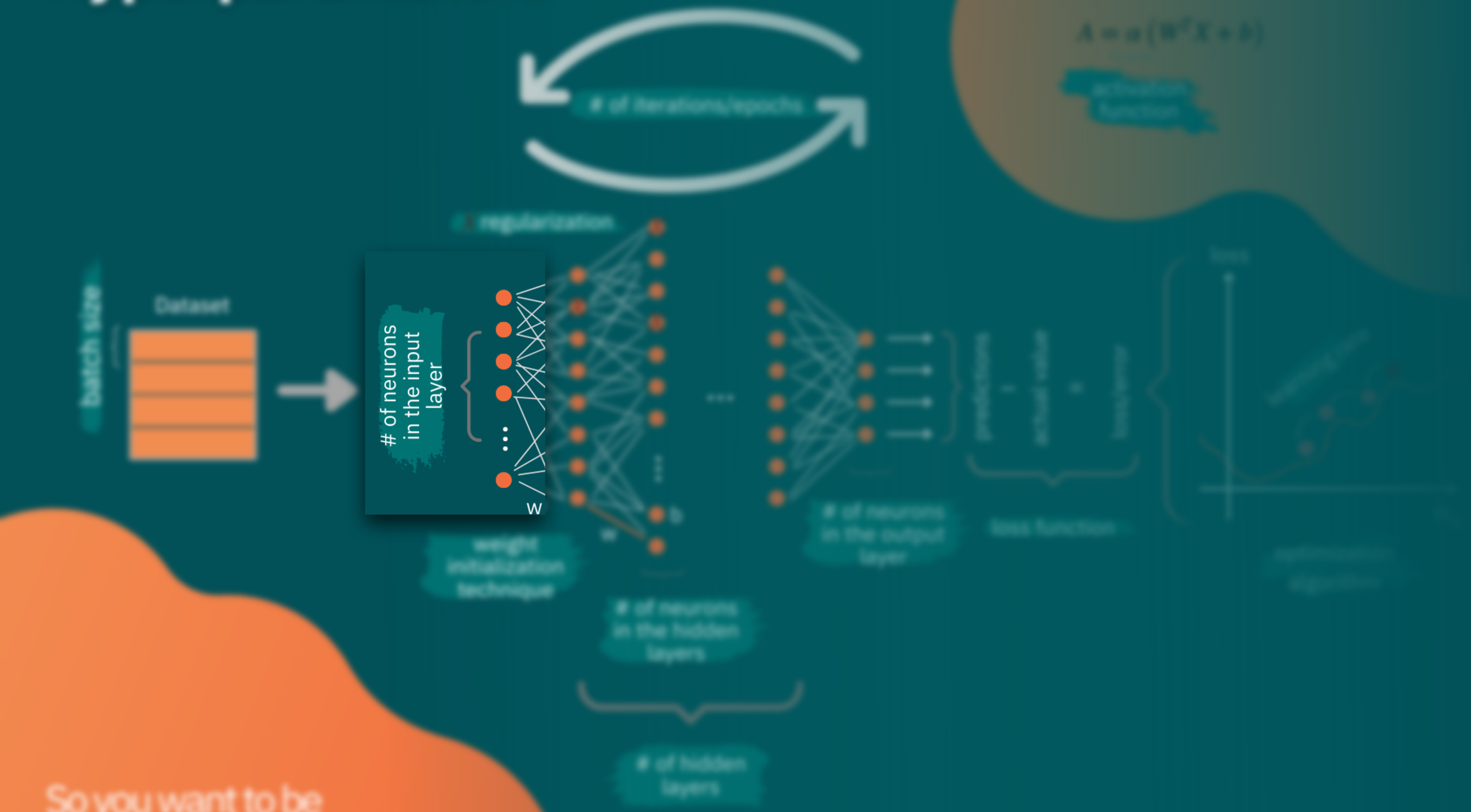
Pre-determined hyperparameters

- Number of neurons in input layer
- Number of neurons in output layer

Hyperparameters that need tuning

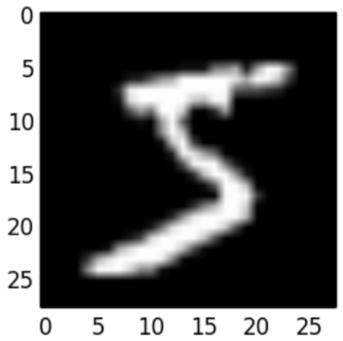
- Number of hidden layers
- Number of neurons in hidden layers
- optimizer
- learning rate
- activation functions
- batch size
- number of iterations
- Loss function
- *Weight initialization technique*
- *Regularization / no regularization (+which reg algorithm)*

Neural Networks Hyperparameters

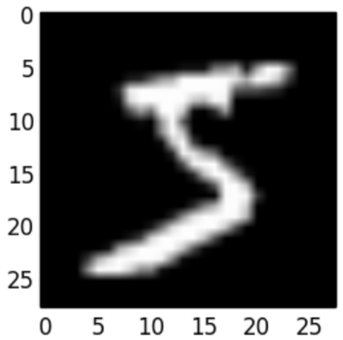


So you want to be a data scientist?

Number of neurons in input layer are determined based on input size

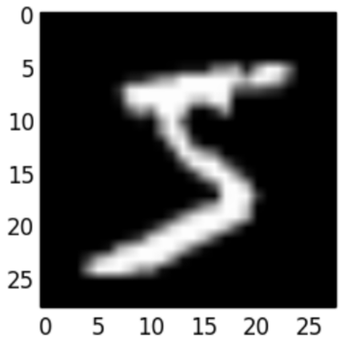


Number of neurons in input layer are determined based on input size



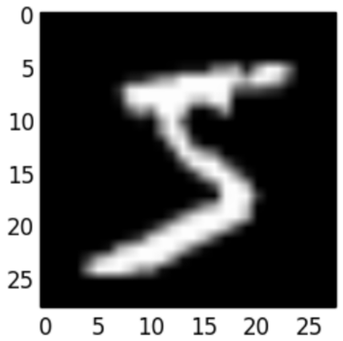
28 x 28

Number of neurons in input layer are determined based on input size



28 x 28 = 784 input neurons

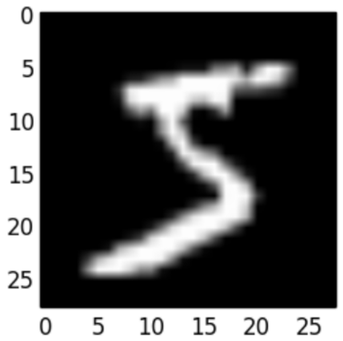
Number of neurons in input layer are determined based on input size



28 x 28 = 784 input neurons

# Number	A Digimon	A Stage	A Type
1	Kuramon	Baby	Free
2	Pabumon	Baby	Free
3	Punimon	Baby	Free
4	Botamon	Baby	Free
5	Poyomon	Baby	Free
6	Koromon	In-Training	Free
7	Tanemon	In-Training	Free
8	Tsunomon	In-Training	Free
9	Tsumemon	In-Training	Free

Number of neurons in input layer are determined based on input size

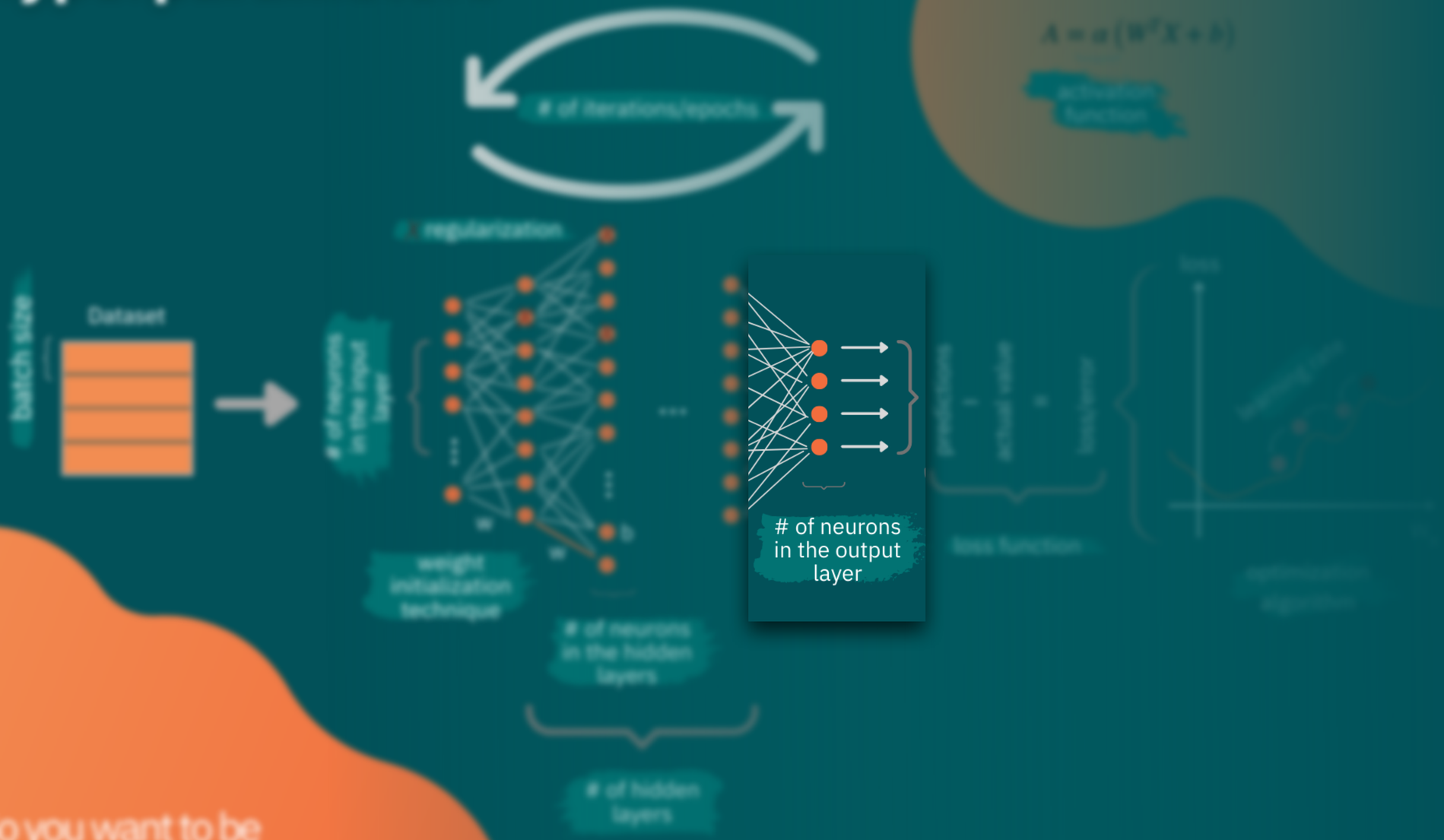


28 x 28 = 784 input neurons

# Number	A Digimon	A Stage	A Type
1	Kuramon	Baby	Free
2	Pabumon	Baby	Free
3	Punimon	Baby	Free
4	Botamon	Baby	Free
5	Poyomon	Baby	Free
6	Koromon	In-Training	Free
7	Tanemon	In-Training	Free
8	Tsunomon	In-Training	Free
9	Tsumemon	In-Training	Free

4 input neurons

Neural Networks Hyperparameters



So you want to be a data scientist?

Number of neurons in output layer are determined based on the output type

Male / female

1 neuron, (0=female, 1=male)

2 neurons (first one=female, second one=male)

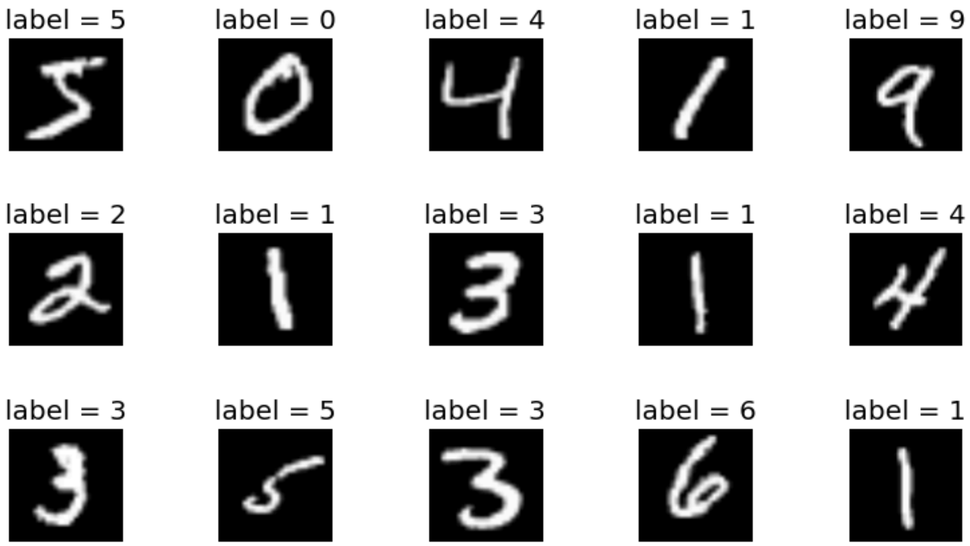
Number of neurons in output layer are determined based on the output type

Male / female → 1 neuron, (0=female, 1=male)

2 neurons (first one=female, second one=male)

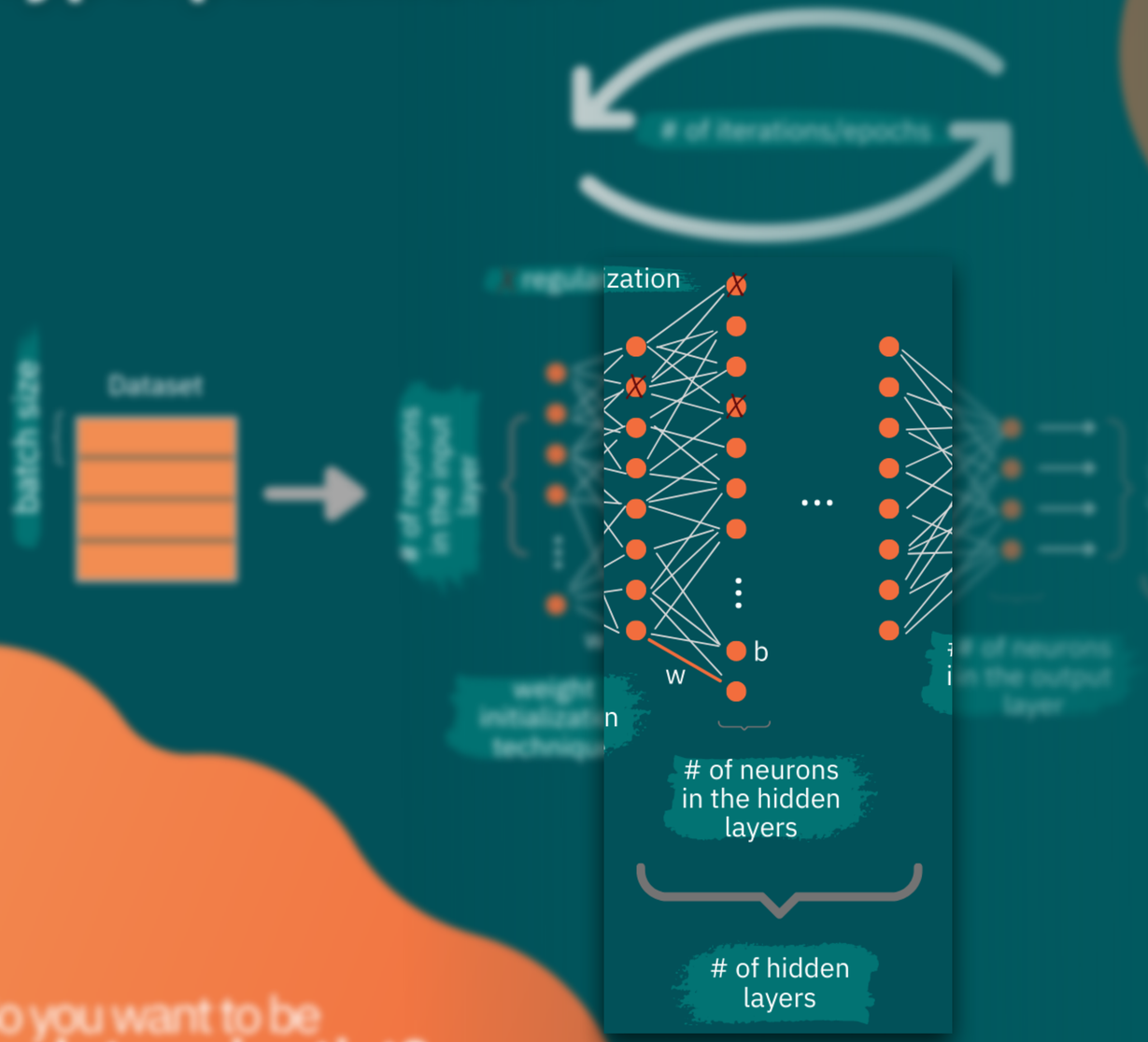
0-200 → Single neuron with the appropriate activation function

Number of neurons in output layer are determined based on the output type



10 output neurons

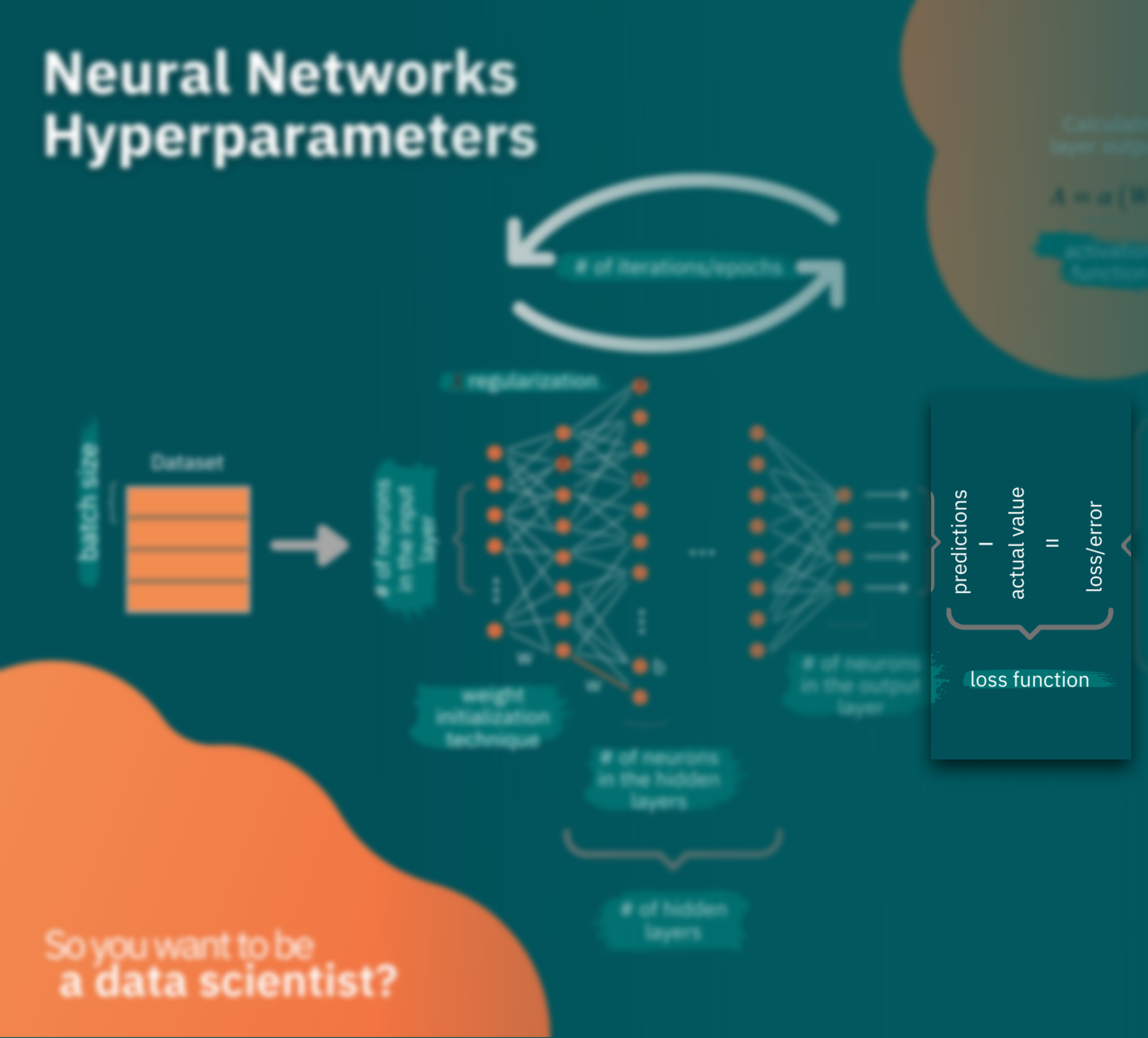
Neural Networks Hyperparameters



Hidden layers

- It’s better you have deeper network than wider network
- Layers learn features at different complexity
- 1-2 layers should be enough for many problems
- Start with 2 and work your way up

So you want to be a data scientist?

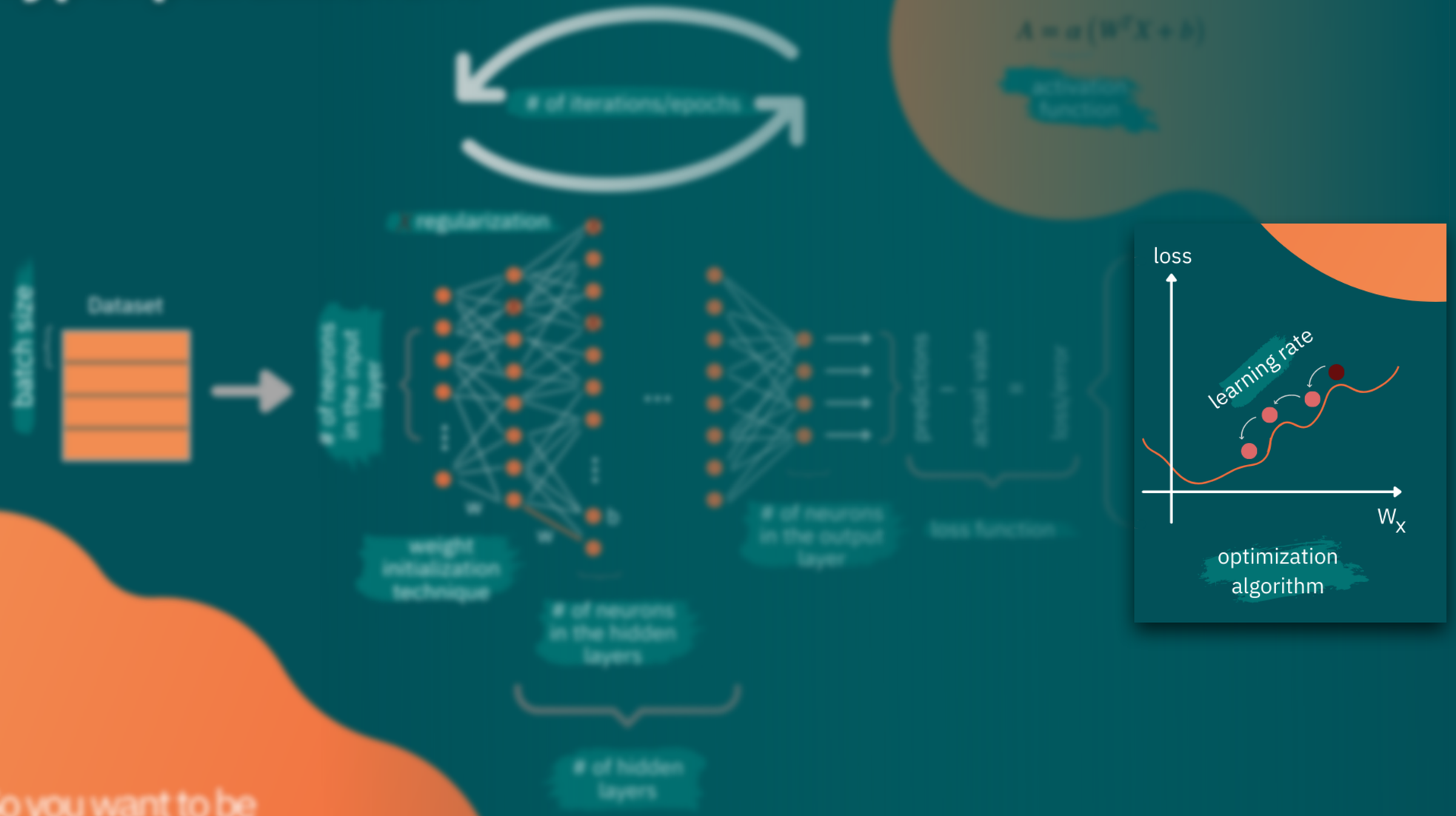


Loss function

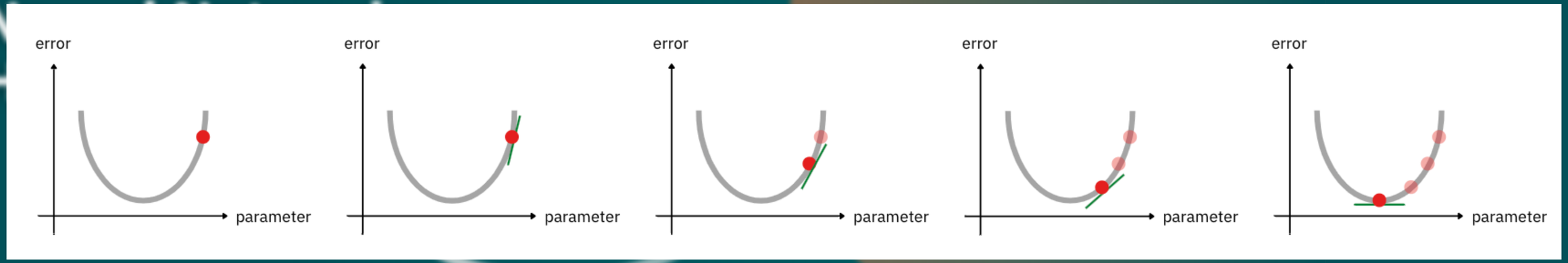
- How the error is calculated
- Some options are:
  - MSE
  - MAE
  - Binary cross-entropy
  - Sparse categorical cross-entropy

Refer to course summary

Neural Networks Hyperparameters



So you want to be a data scientist?

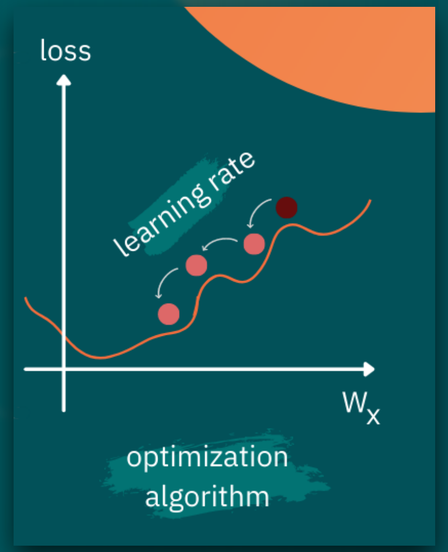


Choosing an optimizer

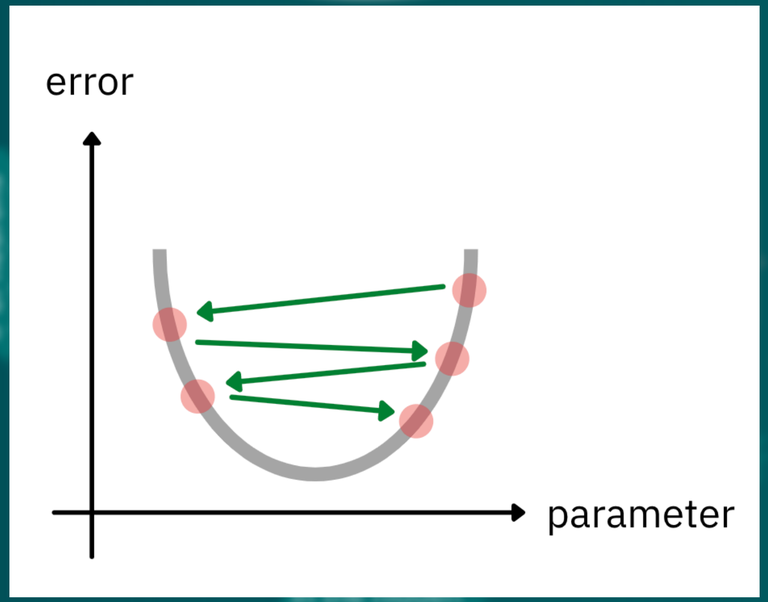
How we decide to update the weights and biases

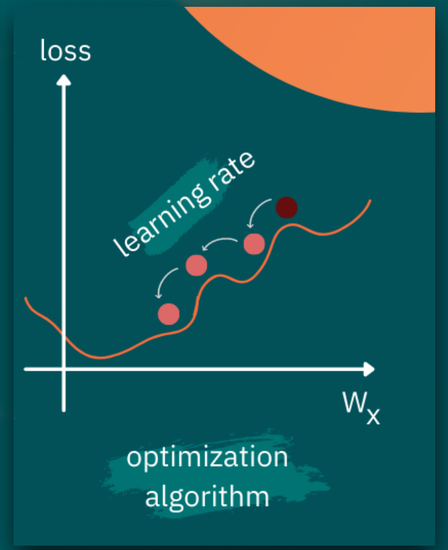
Some other optimizers:

- SGD
- RMSProp
- Adam



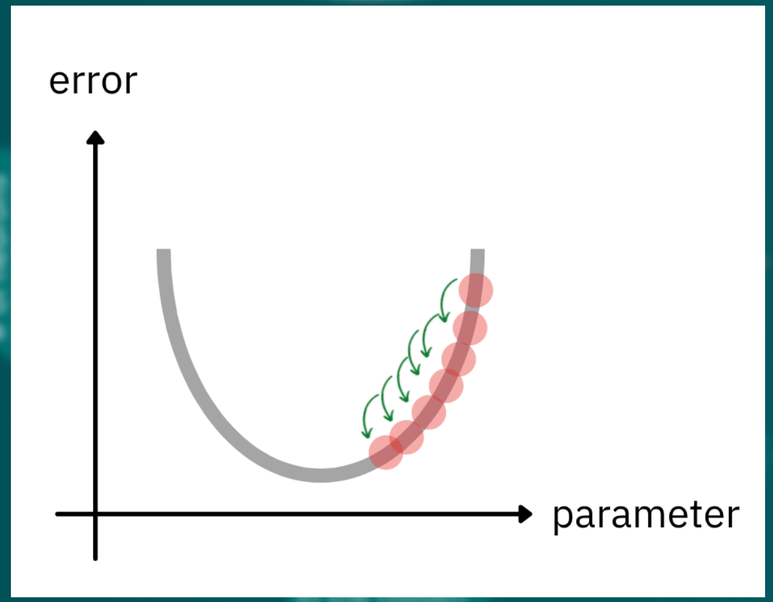
Neural Networks Hyperparameters

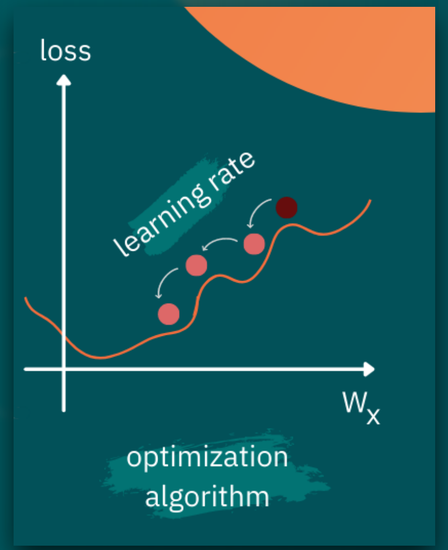




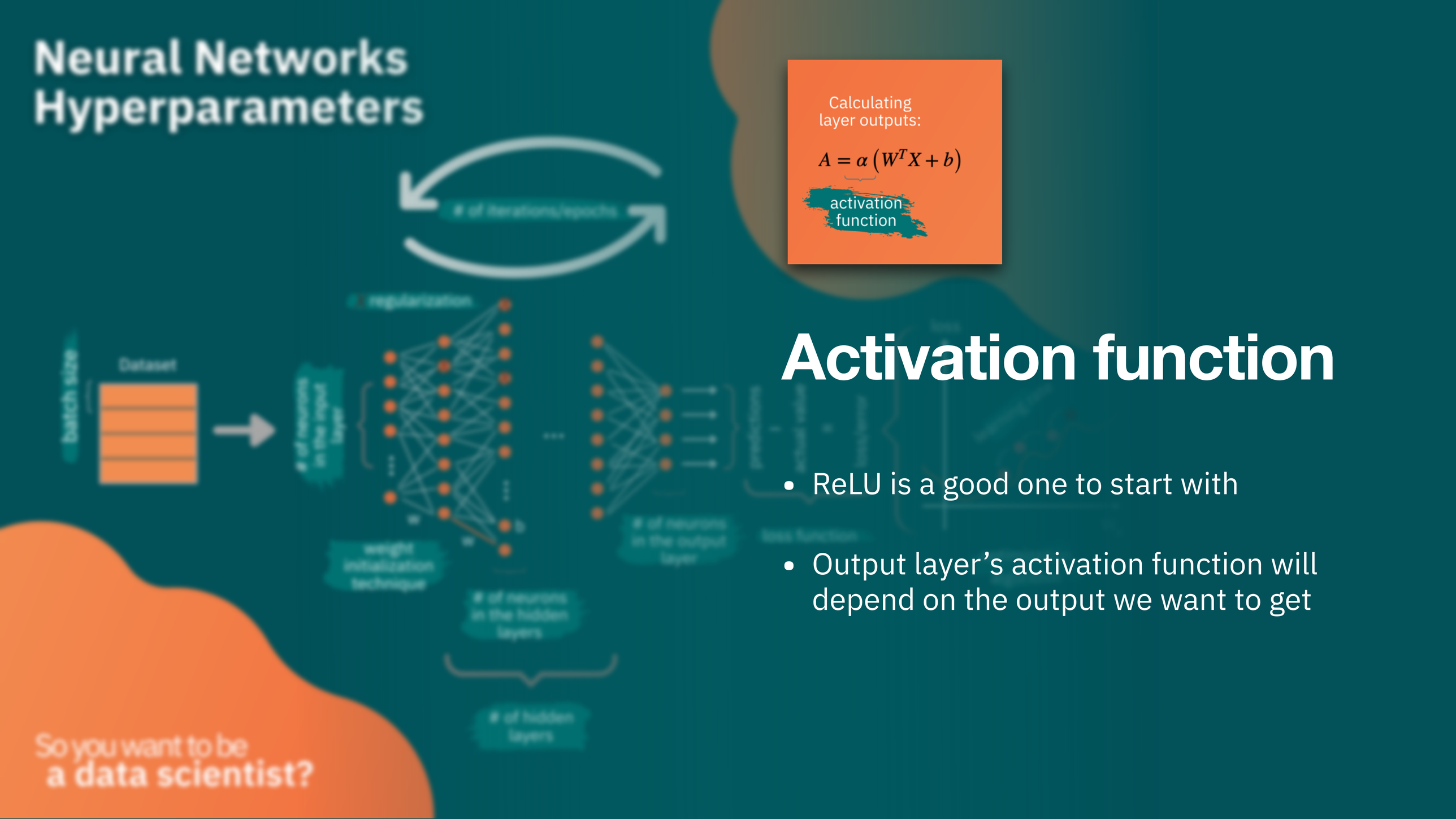
So you want to be a data scientist?

Neural Networks Hyperparameters





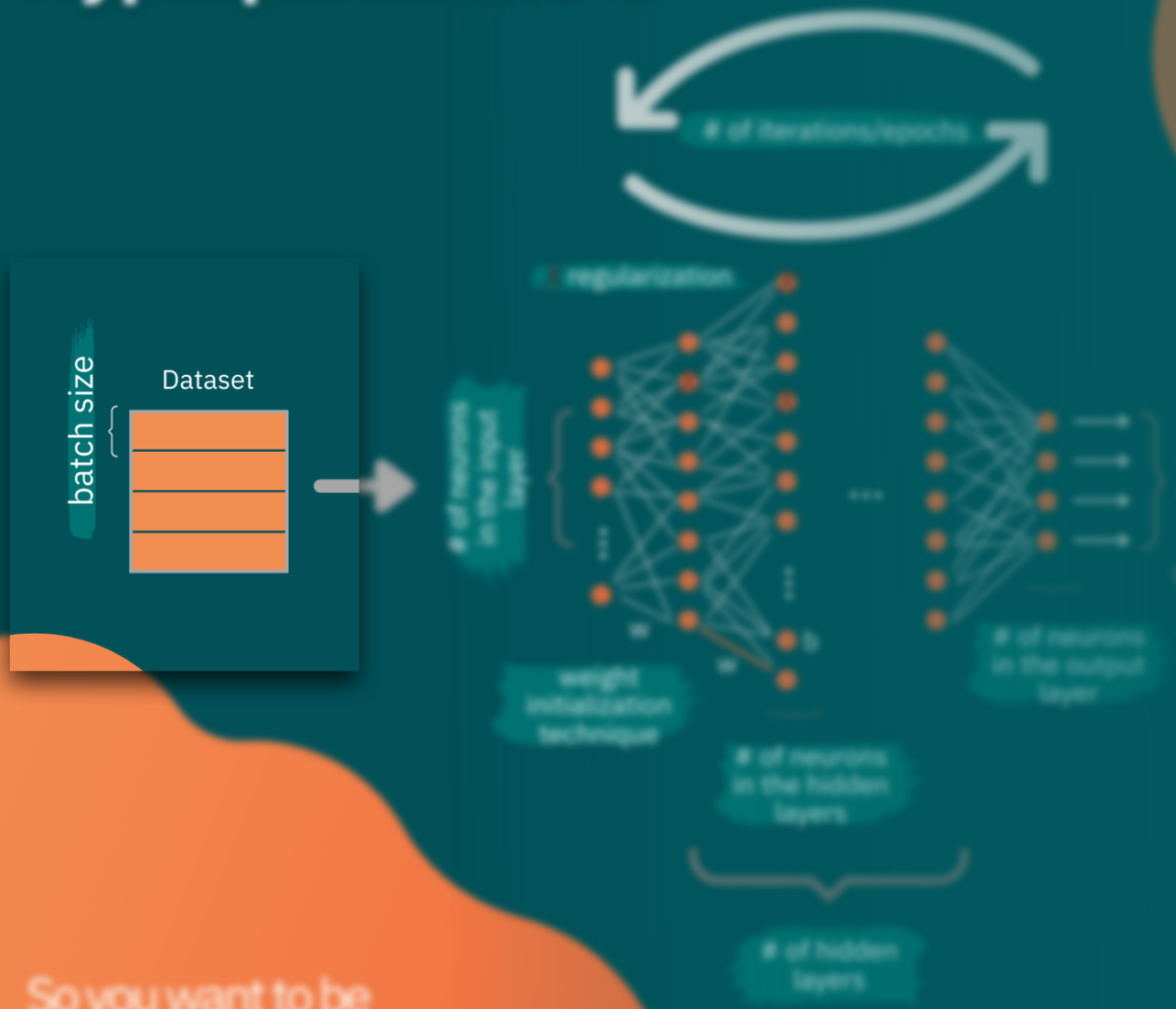
So you want to be a data scientist?



Activation function

- ReLU is a good one to start with
- Output layer's activation function will depend on the output we want to get

Neural Networks Hyperparameters



Batch size

- Batching means separating data into smaller pieces
- Processors can run through them more quickly
- Helps with generalization

You will hear it used with Gradient Descent:

- Batch GD
- Mini-batch GD
- Stochastic GD

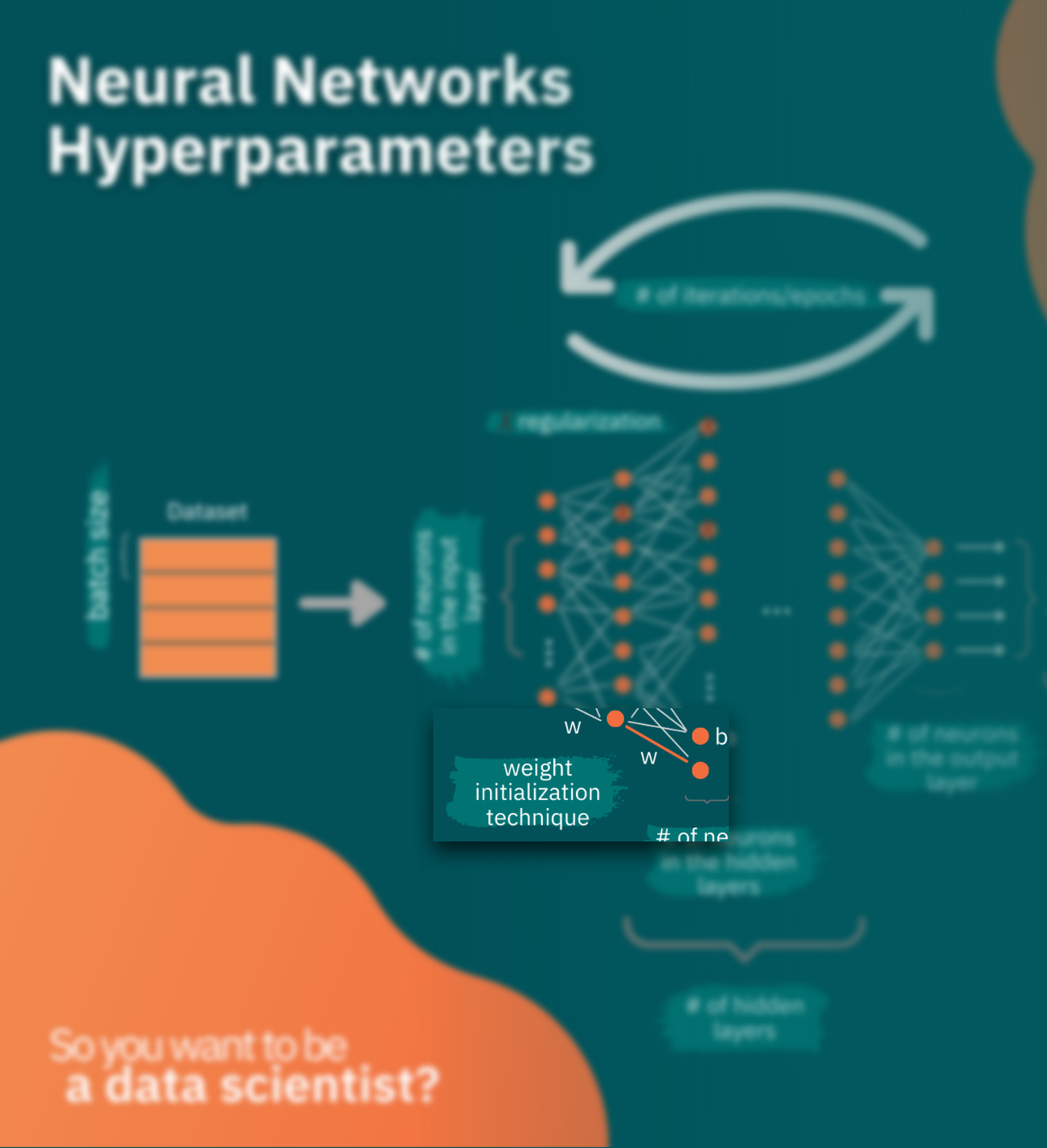
So you want to be a data scientist?

Neural Networks Hyperparameters

of iterations (epochs)

- You can decide this based on your model's performance
- Early stopping is a way to determine the optimal number

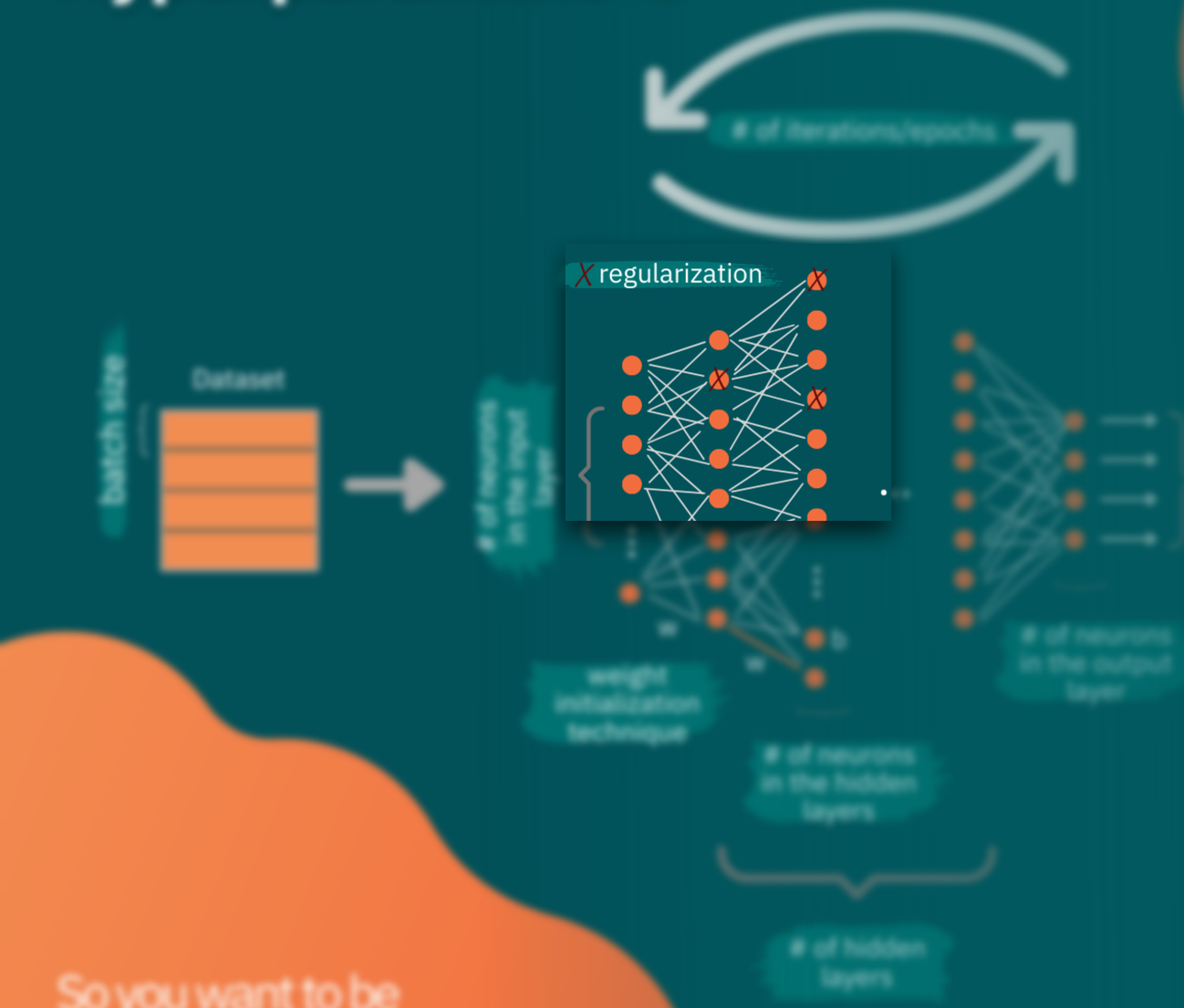
So you want to be a data scientist?



Weight initialization

- Weights must be initialized randomly
- Bias's are all set to zero
- Changing initialization method to have a certain variance of weights at initial state can help with unstable gradients problem

Neural Networks Hyperparameters



Regularization

- Makes a simpler network
- We use it to deal with overfitting of Neural Networks
- Has its own hyperparemeters

So you want to be a data scientist?