

# Quiz week 4: artificial neural networks

# Q1

Artificial neural networks (ANN) are used for both supervised and unsupervised tasks. Which of the following statements are correct?

- 1. A feedforward network can accurately approximate any continuous function**
- 2. ANN can model non-linear relationships**
- 3. ANN have input, hidden and output layers**
4. ANN's predictions are easy to understand

# Q1

Artificial neural networks (ANN) are used for both supervised and unsupervised tasks. Which of the following statements are correct?

**1. A feedforward network can accurately approximate any continuous function**

Universal approximation theorem: a feedforward network can accurately approximate any *continuous* function from one finite dimensional space to another, *given enough hidden units* (Hornik et al. 1989, Cybenko 1989).

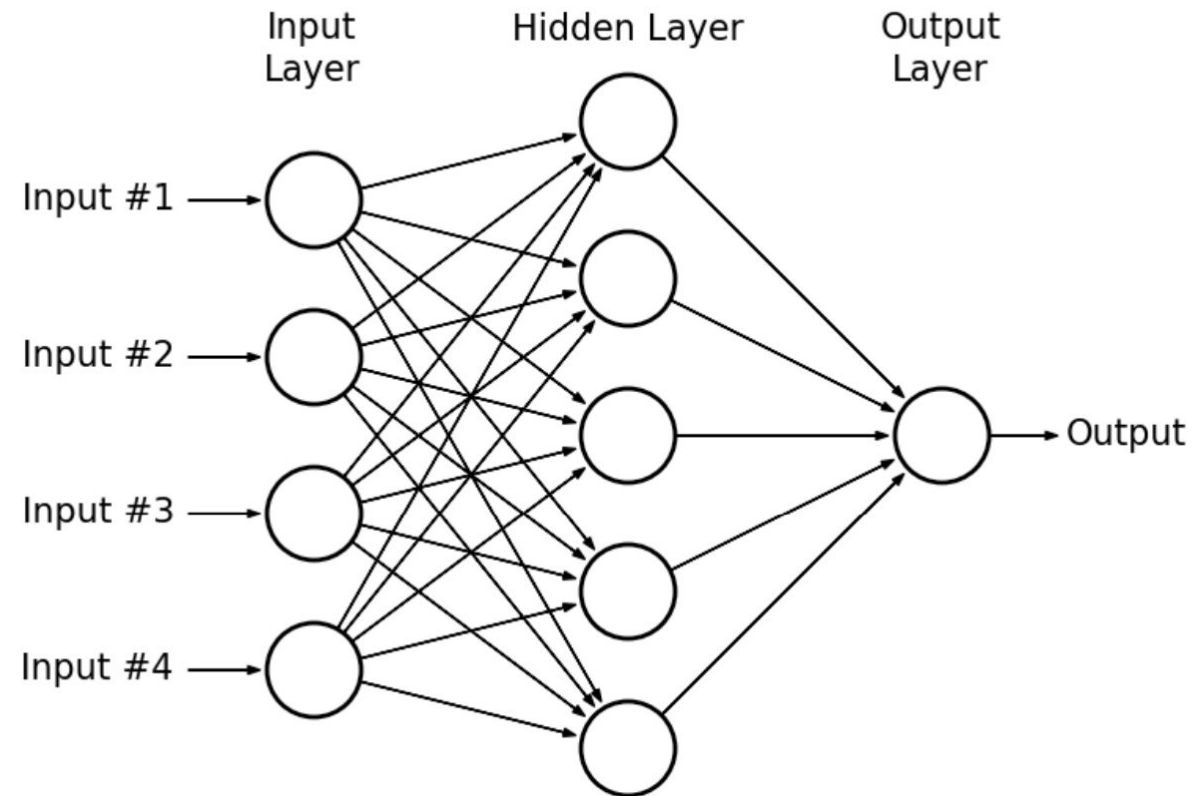
# Q1

Artificial neural networks (ANN) are used for both supervised and unsupervised tasks.  
Which of the following statements are correct?

## 2. ANN can model non-linear relationships

To model non-linear relationships ANN needs:

1. Hidden layers which are connected to all input's features



# Q1

Artificial neural networks (ANN) are used for both supervised and unsupervised tasks.  
Which of the following statements are correct?

## 2. ANN can model non-linear relationships

To model non-linear relationships ANN needs:

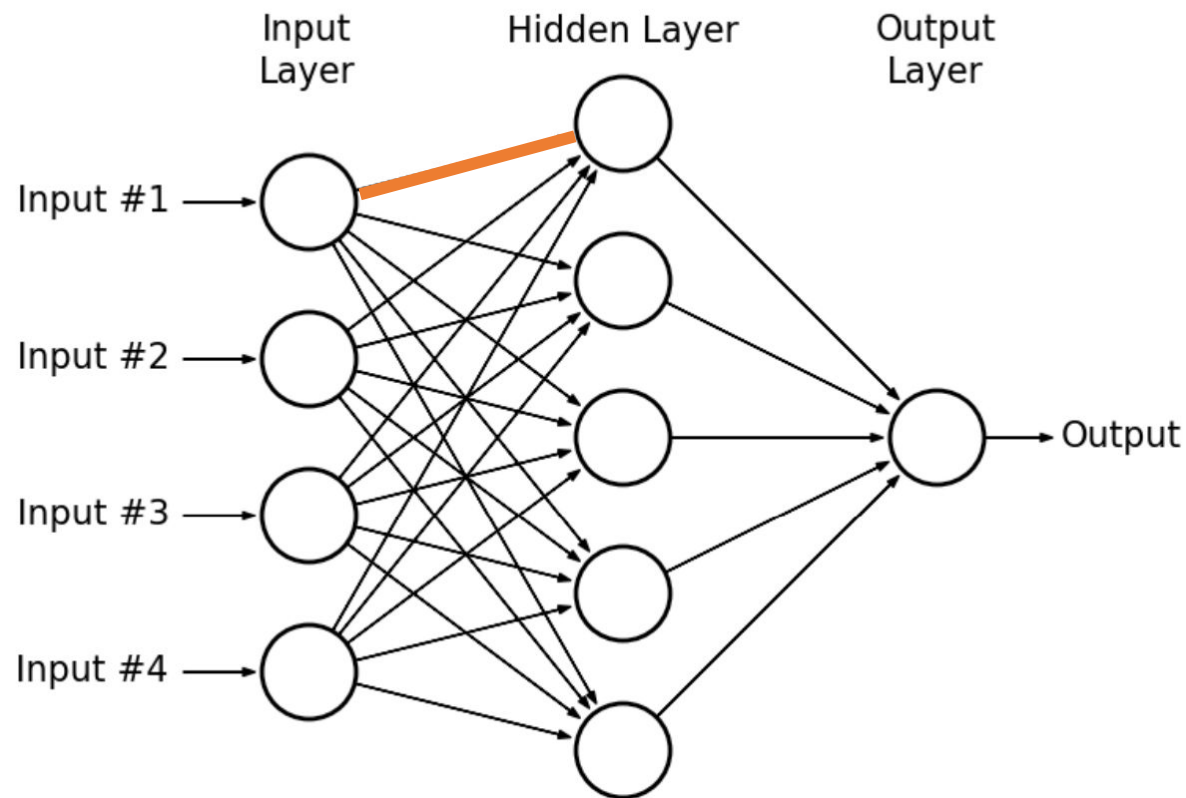
1. Hidden layers which are connected to all input's features

$$h_1 = \underline{W_{1,1}^{(1)}}x_1 + W_{1,2}^{(1)}x_2 + \cdots + W_{1,d}^{(1)}x_d + b_1$$

$$h_2 = W_{2,1}^{(1)}x_1 + W_{2,2}^{(1)}x_2 + \cdots + W_{2,d}^{(1)}x_d + b_1$$

$\vdots$

$$h_{d_1} = W_{d_1,1}^{(1)}x_1 + W_{d_1,2}^{(1)}x_2 + \cdots + W_{d_1,d}^{(1)}x_d + b_1$$



# Q1

Artificial neural networks (ANN) are used for both supervised and unsupervised tasks.  
Which of the following statements are correct?

## 2. ANN can model non-linear relationships

To model non-linear relationships ANN needs:

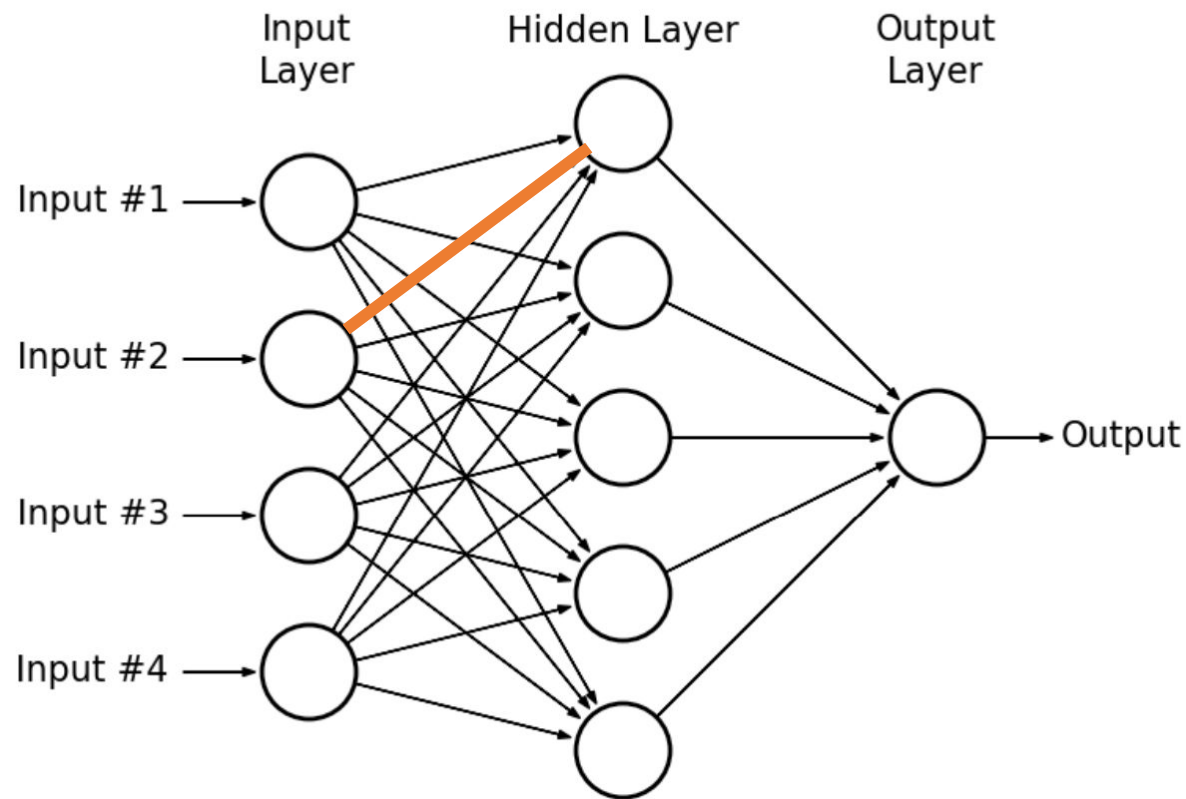
1. Hidden layers which are connected to all input's features

$$h_1 = W_{1,1}^{(1)}x_1 + \underline{W_{1,2}^{(1)}}x_2 + \cdots + W_{1,d}^{(1)}x_d + b_1$$

$$h_2 = W_{2,1}^{(1)}x_1 + W_{2,2}^{(1)}x_2 + \cdots + W_{2,d}^{(1)}x_d + b_2$$

$\vdots$

$$h_{d_1} = W_{d_1,1}^{(1)}x_1 + W_{d_1,2}^{(1)}x_2 + \cdots + W_{d_1,d}^{(1)}x_d + b_{d_1}$$



# Q1

Artificial neural networks (ANN) are used for both supervised and unsupervised tasks.  
Which of the following statements are correct?

## 2. ANN can model non-linear relationships

To model non-linear relationships ANN needs:

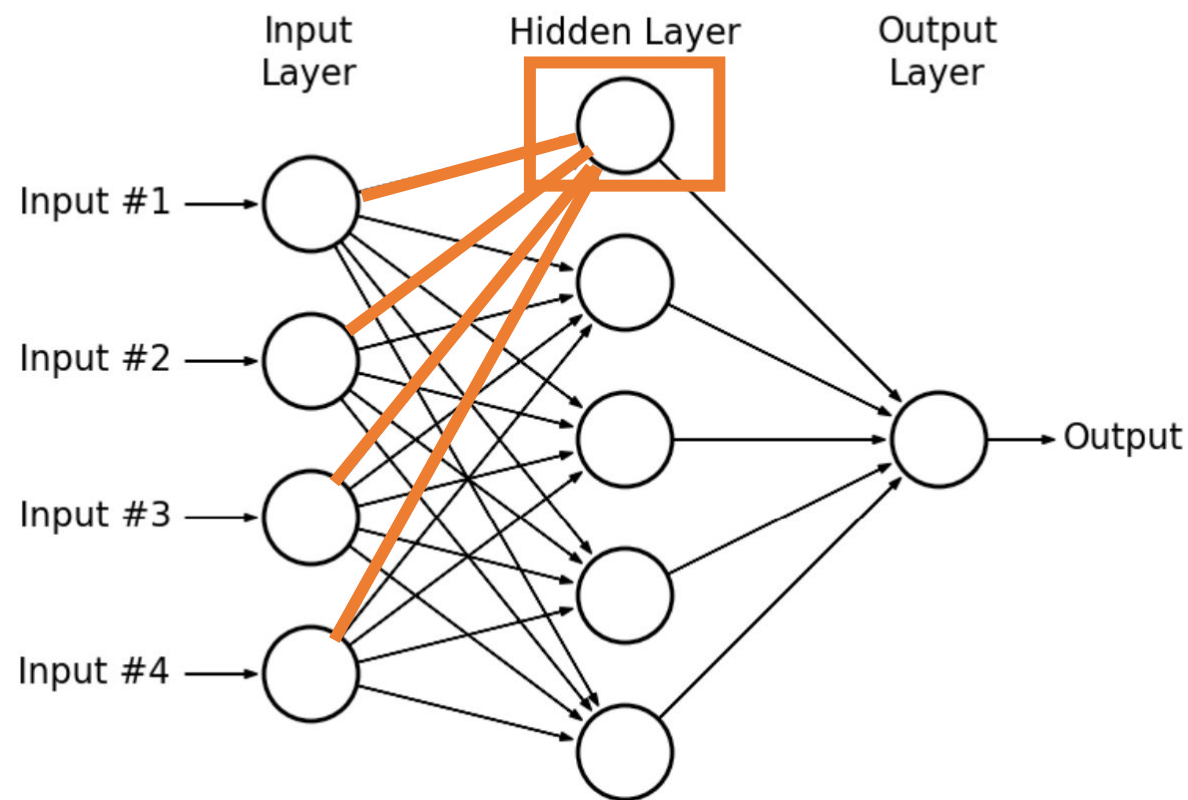
1. Hidden layers which are connected to all input's features

$$h_1 = W_{1,1}^{(1)}x_1 + W_{1,2}^{(1)}x_2 + \cdots + W_{1,d}^{(1)}x_d + b_1$$

$$h_2 = W_{2,1}^{(1)}x_1 + W_{2,2}^{(1)}x_2 + \cdots + W_{2,d}^{(1)}x_d + b_1$$

$\vdots$

$$h_{d_1} = W_{d_1,1}^{(1)}x_1 + W_{d_1,2}^{(1)}x_2 + \cdots + W_{d_1,d}^{(1)}x_d + b_1$$



# Q1

Artificial neural networks (ANN) are used for both supervised and unsupervised tasks.  
Which of the following statements are correct?

## 2. ANN can model non-linear relationships

To model non-linear relationships ANN needs:

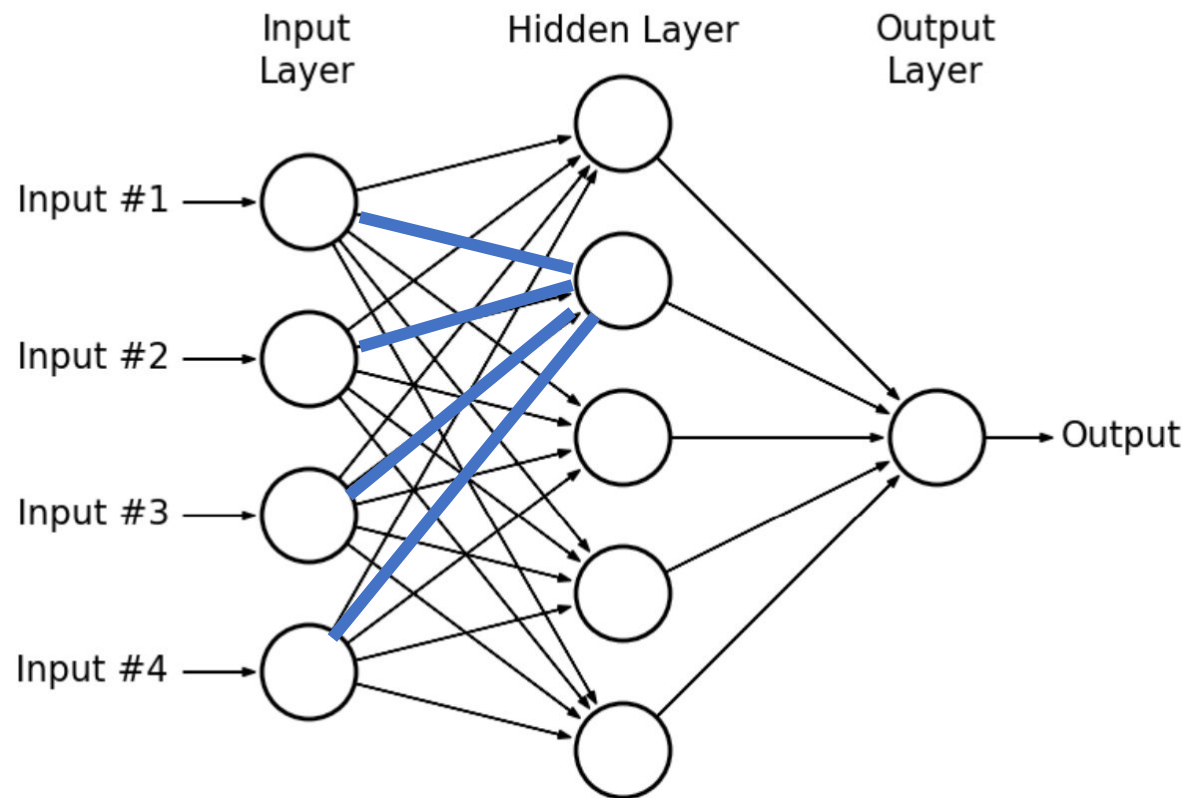
1. Hidden layers which are connected to all input's features

$$h_1 = W_{1,1}^{(1)}x_1 + W_{1,2}^{(1)}x_2 + \cdots + W_{1,d}^{(1)}x_d + b_1$$

$$h_2 = W_{2,1}^{(1)}x_1 + W_{2,2}^{(1)}x_2 + \cdots + W_{2,d}^{(1)}x_d + b_1$$

$\vdots$

$$h_{d_1} = W_{d_1,1}^{(1)}x_1 + W_{d_1,2}^{(1)}x_2 + \cdots + W_{d_1,d}^{(1)}x_d + b_1$$





# Q1

Artificial neural networks (ANN) are used for both supervised and unsupervised tasks.  
Which of the following statements are correct?

## 2. ANN can model non-linear relationships

To model non-linear relationships ANN needs:

1. Hidden layers which are connected to all input's features
2. A non-linear activation function

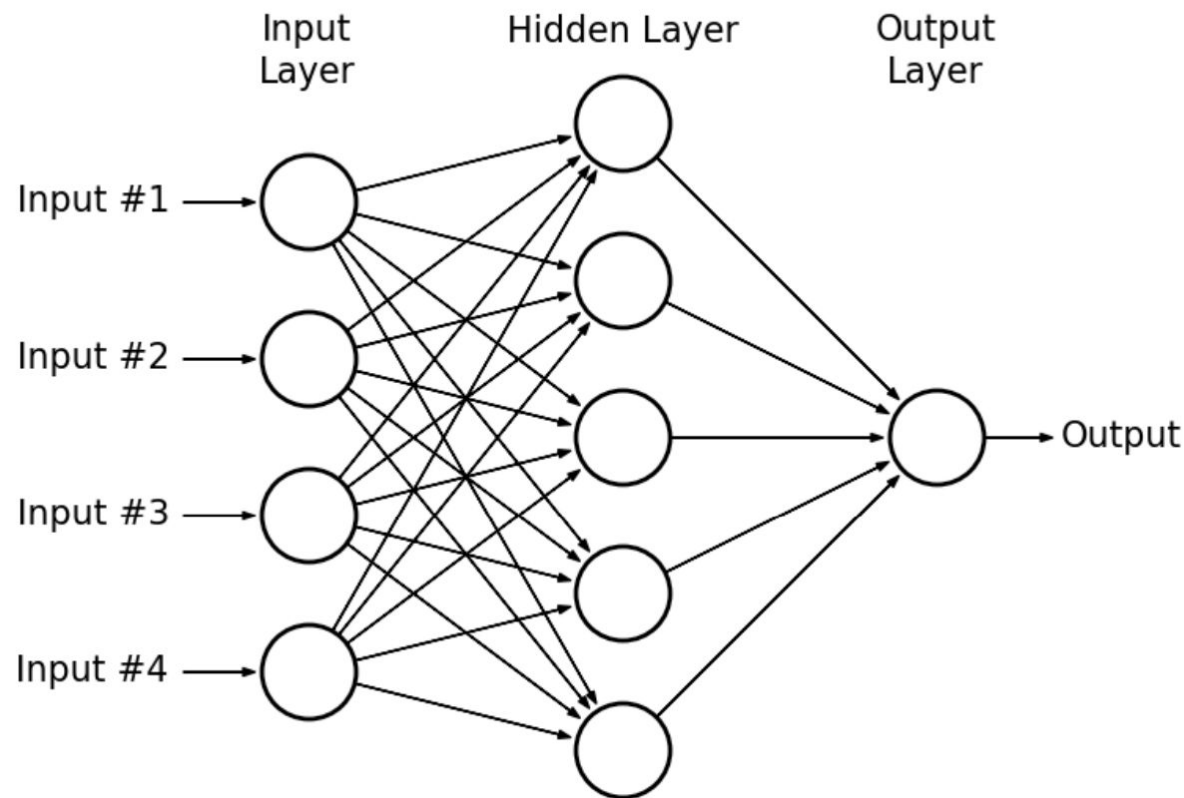
This is a **linear** combination of interconnected layers

$$h_1 = W_{1,1}^{(1)}x_1 + W_{1,2}^{(1)}x_2 + \cdots + W_{1,d}^{(1)}x_d + b_1$$

$$h_2 = W_{2,1}^{(1)}x_1 + W_{2,2}^{(1)}x_2 + \cdots + W_{2,d}^{(1)}x_d + b_1$$

$\vdots$

$$h_{d_1} = W_{d_1,1}^{(1)}x_1 + W_{d_1,2}^{(1)}x_2 + \cdots + W_{d_1,d}^{(1)}x_d + b_1$$



# Q1

Artificial neural networks (ANN) are used for both supervised and unsupervised tasks.  
Which of the following statements are correct?

## 2. ANN can model non-linear relationships

To model non-linear relationships ANN needs:

1. Hidden layers which are connected to all input's features
2. A non-linear activation function

This is a **linear** combination of interconnected layers

$$h_1 = W_{1,1}^{(1)}x_1 + W_{1,2}^{(1)}x_2 + \cdots + W_{1,d}^{(1)}x_d + b_1$$

$$h_2 = W_{2,1}^{(1)}x_1 + W_{2,2}^{(1)}x_2 + \cdots + W_{2,d}^{(1)}x_d + b_1$$

$\vdots$

$$h_{d_1} = W_{d_1,1}^{(1)}x_1 + W_{d_1,2}^{(1)}x_2 + \cdots + W_{d_1,d}^{(1)}x_d + b_1$$



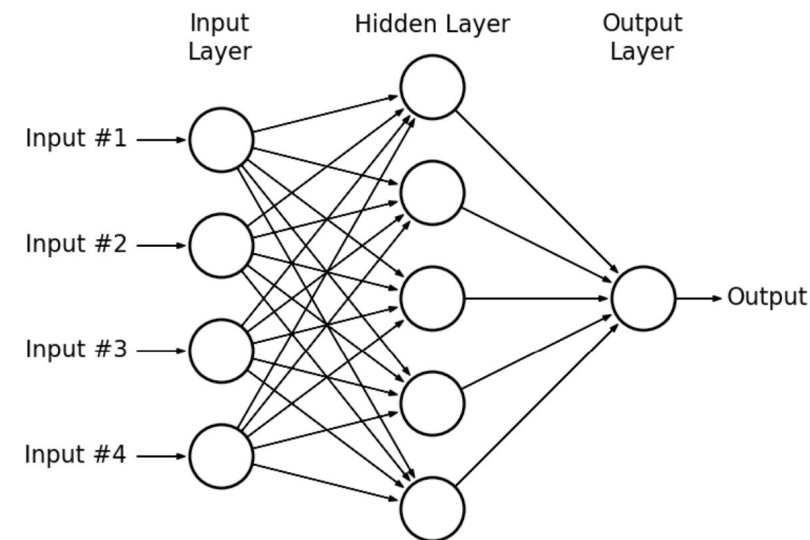
Apply a non-linear function  $\sigma$  to obtain the predictions from the model

$$f(\mathbf{x}; \theta) = \sigma(\mathbf{W}_2 \sigma(\mathbf{W}_1 \mathbf{x} + b_1) + b_2)$$

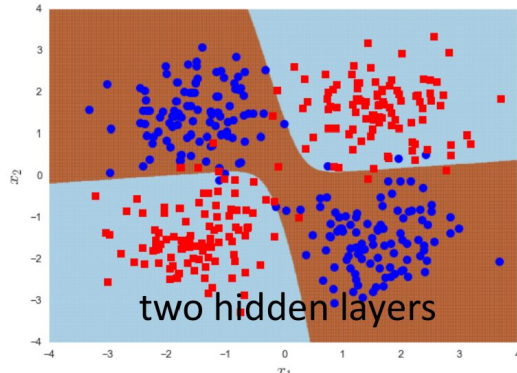
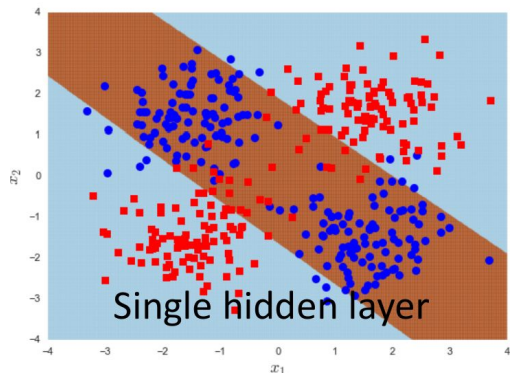
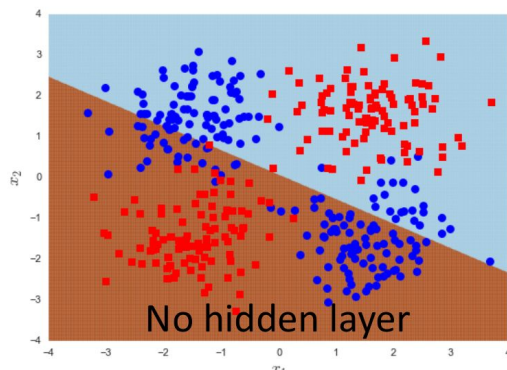
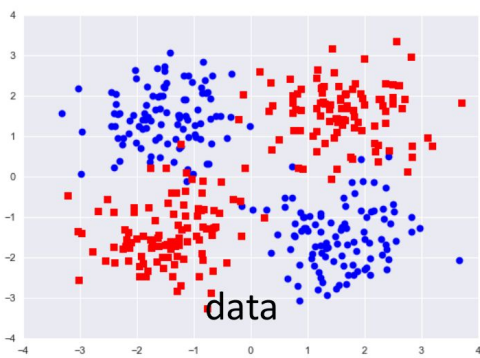
# Q1

Artificial neural networks (ANN) are used for both supervised and unsupervised tasks. Which of the following statements are correct?

**3. ANN have input, hidden and output layers**



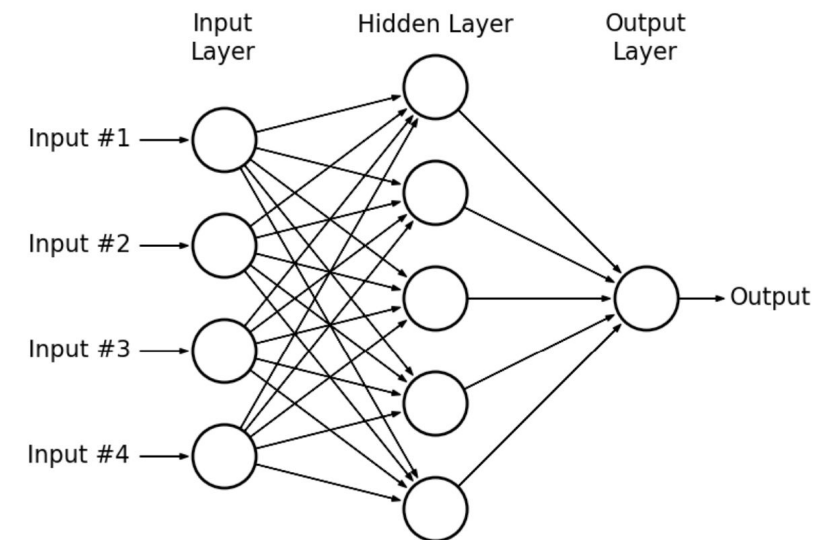
Q: Increasing the number of hidden layers (depth) allows more complicated decision boundaries, why?



# Q1

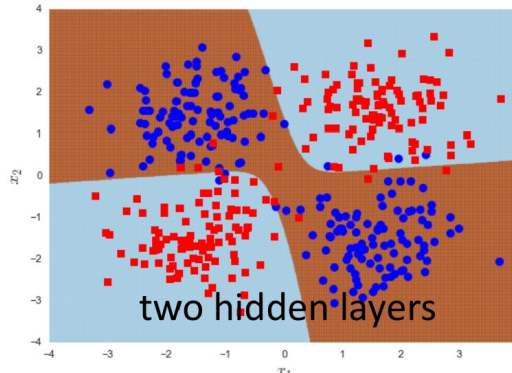
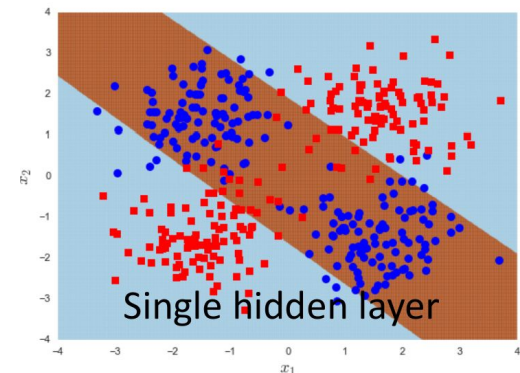
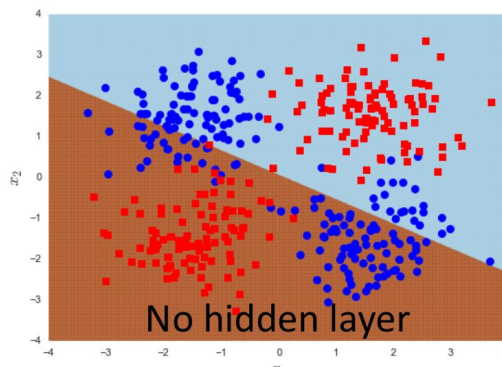
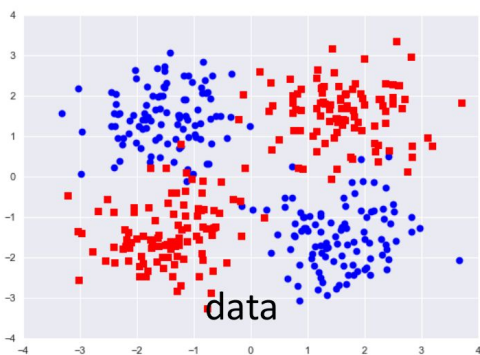
Artificial neural networks (ANN) are used for both supervised and unsupervised tasks. Which of the following statements are correct?

**3. ANN have input, hidden and output layers**



Q: Increasing the number of hidden layers (depth) allows more complicated decision boundaries, why?

A: Increasing depth implies more parameters (weights,  $W$ ) = higher number of connections between the input (features) and the output (predictions)



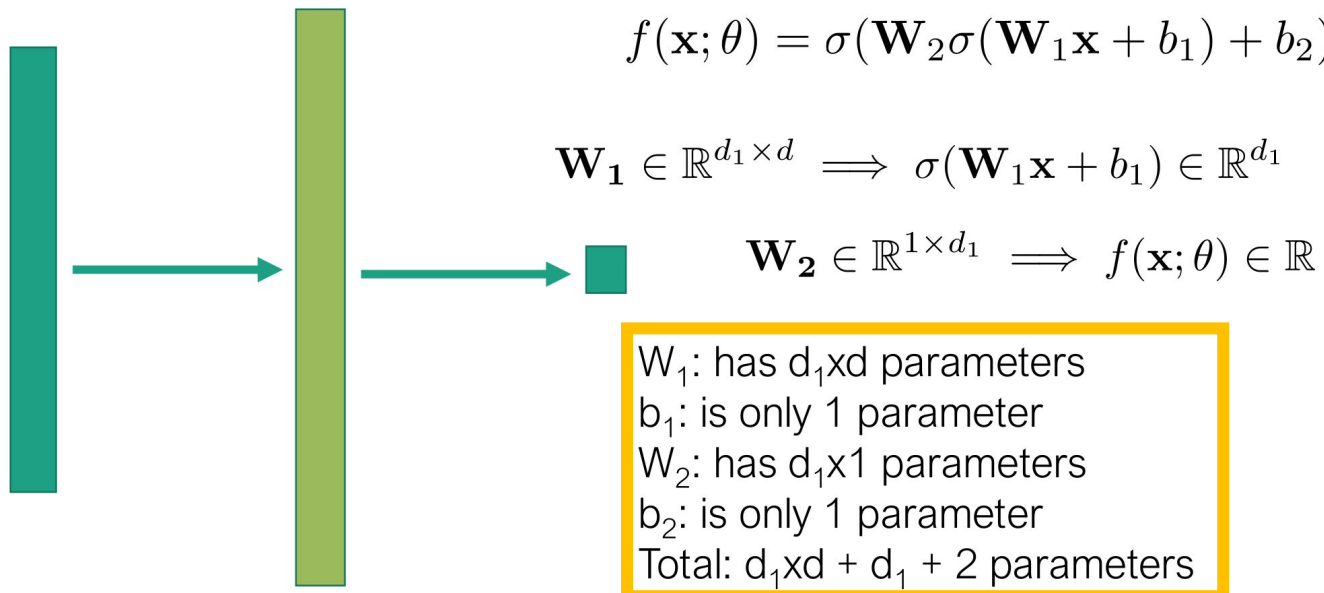
# Q1

Artificial neural networks (ANN) are used for both supervised and unsupervised tasks. Which of the following statements are correct?

4. ANN's predictions are easy to understand

With many interconnected **parameters** it's hard to understand the contributions to the predictions from the features present in the data

**Number of parameters increased**



# Q2

Which of the following statements are true about training ANN?

- 1. Training ANN is a two-step process**
- 2. Weighted sum of linear combinations of input passing through a non-linear activation function**
3. ANN don't have hyper-parameters
- 4. During the learning process, ANN perform optimisation to find the optimal values of the weights**



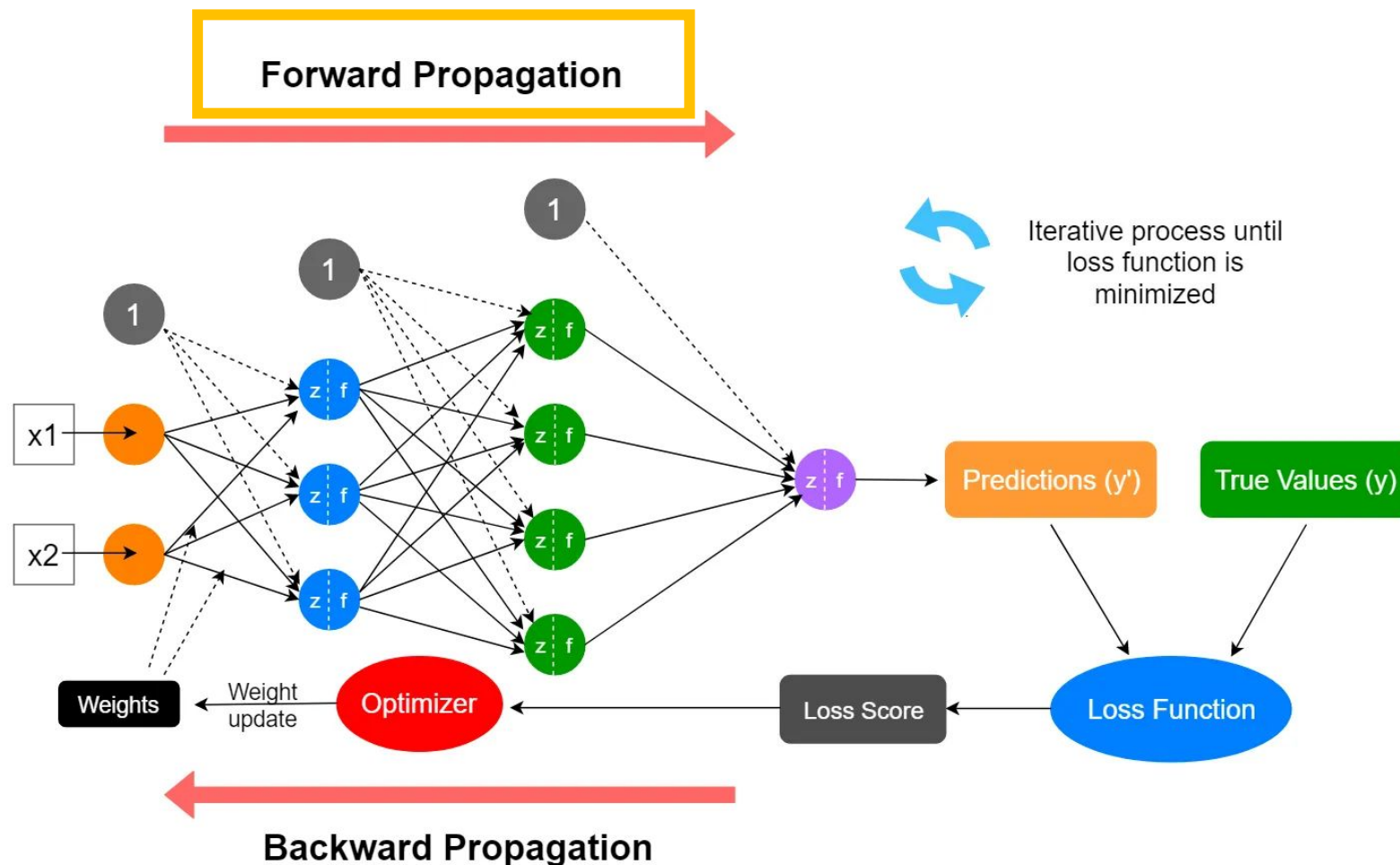
# Q2

Which of the following statements are true about training ANN?

## 1. Training ANN is a two-step process

The forward propagation computes the weights

$$\begin{aligned}h_1 &= W_{1,1}^{(1)}x_1 + W_{1,2}^{(1)}x_2 + \dots + W_{1,d}^{(1)}x_d + b_1 \\h_2 &= W_{2,1}^{(1)}x_1 + W_{2,2}^{(1)}x_2 + \dots + W_{2,d}^{(1)}x_d + b_1 \\&\vdots \\h_{d_1} &= W_{d_1,1}^{(1)}x_1 + W_{d_1,2}^{(1)}x_2 + \dots + W_{d_1,d}^{(1)}x_d + b_1 \\f(\mathbf{x}; \theta) &= \sigma(\mathbf{W}_2 \sigma(\mathbf{W}_1 \mathbf{x} + b_1) + b_2)\end{aligned}$$

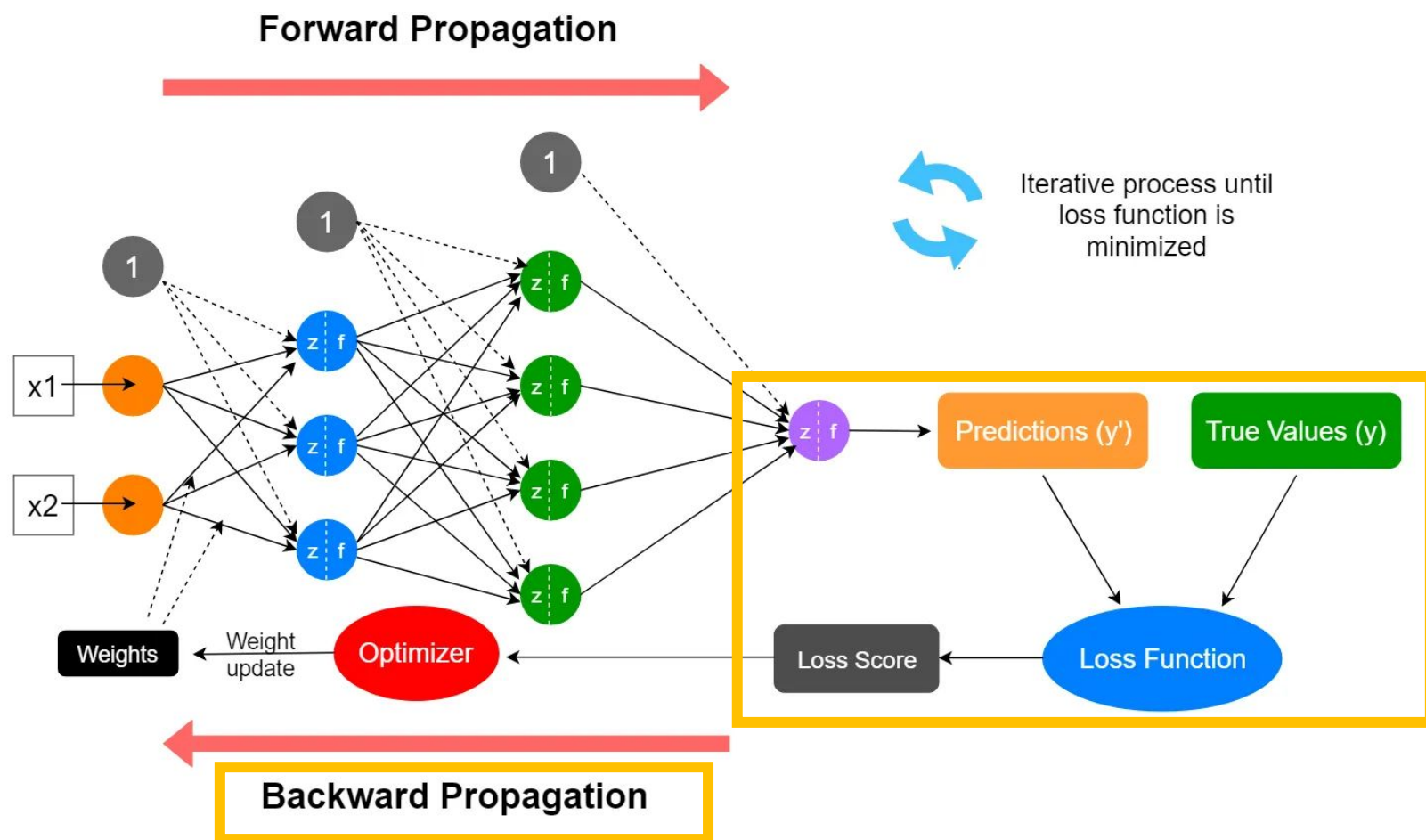


# Q2

Which of the following statements are true about training ANN?

## 1. Training ANN is a two-step process

The backward propagation updates the weights based on the errors from the forward propagation

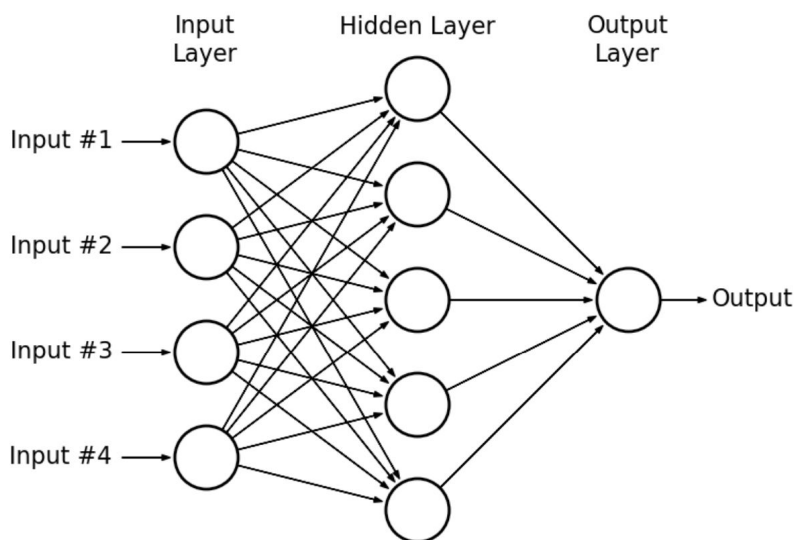




# Q2

Which of the following statements are true about training ANN?

**2. Weighted sum of linear combinations of input passing through a non-linear activation function**



$$\begin{aligned}h_1 &= W_{1,1}^{(1)}x_1 + W_{1,2}^{(1)}x_2 + \cdots + W_{1,d}^{(1)}x_d + b_1 \\h_2 &= W_{2,1}^{(1)}x_1 + W_{2,2}^{(1)}x_2 + \cdots + W_{2,d}^{(1)}x_d + b_1 \\&\vdots\end{aligned}$$

$$h_{d_1} = W_{d_1,1}^{(1)}x_1 + W_{d_1,2}^{(1)}x_2 + \cdots + W_{d_1,d}^{(1)}x_d + b_{d_1}$$



Apply a non-linear function  $\sigma$  to obtain the predictions from the model

$$f(\mathbf{x}; \theta) = \sigma(\mathbf{W}_2 \sigma(\mathbf{W}_1 \mathbf{x} + b_1) + b_2)$$

# Q2

Which of the following statements are true about training ANN?

3. ANN don't have hyper-parameters

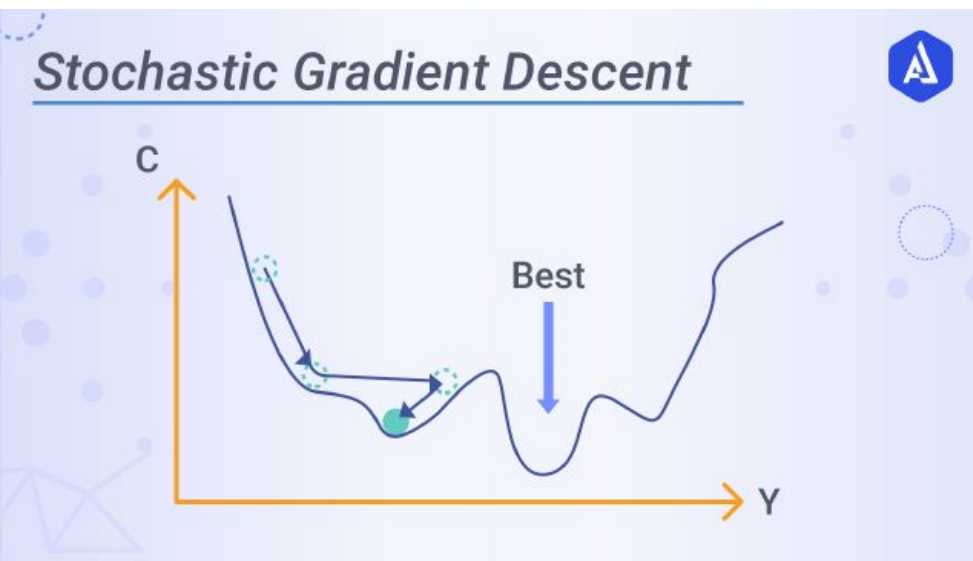
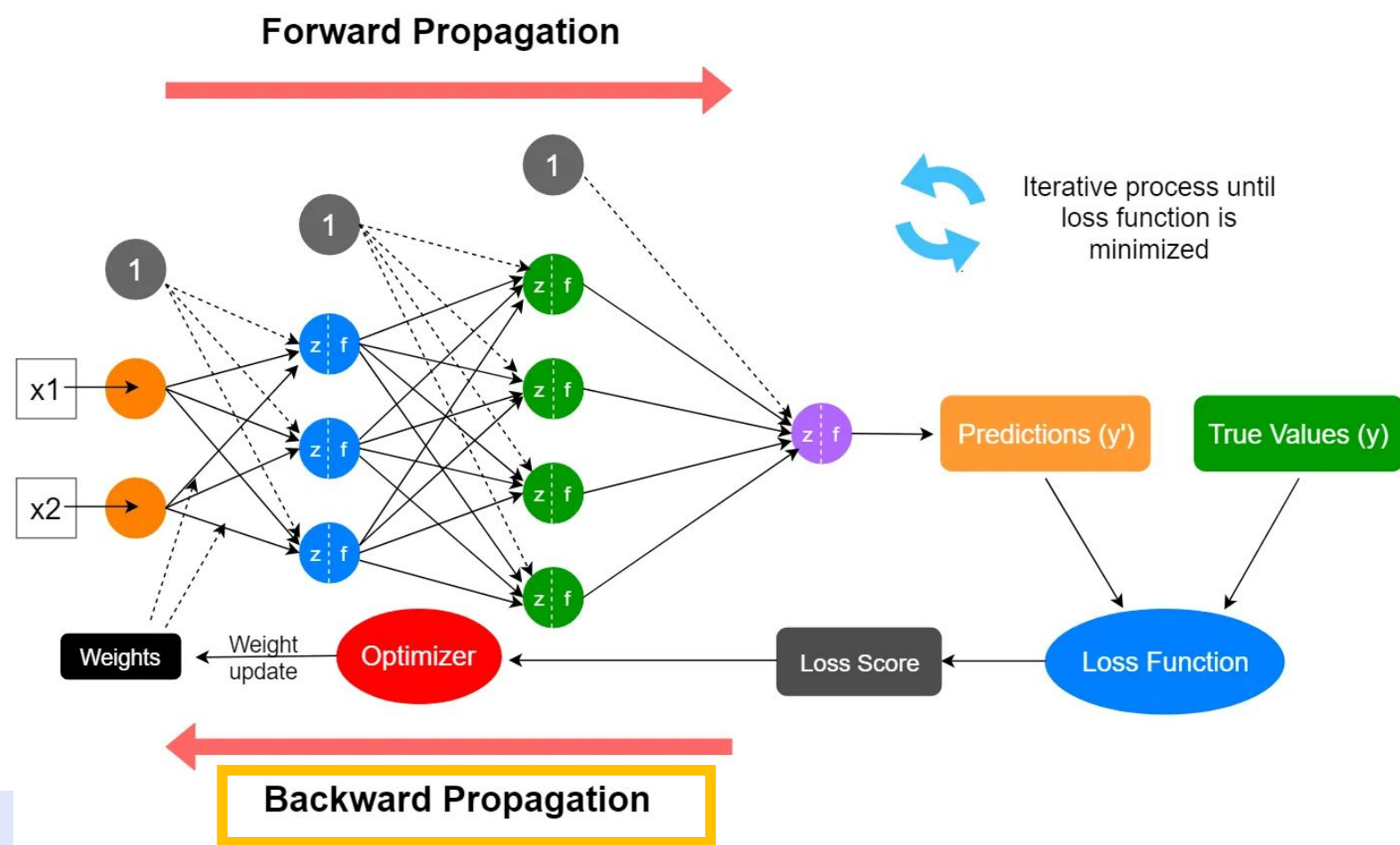
Some hyper-parameters:

- Number of neurons in the hidden layers
- Number of hidden layers
- Activation functions (sigmoid, RELU, tanh)
- ...

## Q2

Which of the following statements are true about training ANN?

**4. During the learning process, ANN perform optimisation to find the optimal values of the weights**



During backward propagation, we use optimisation algorithms such as stochastic gradient descent which optimises the loss function

# Q3

Which of the following are correct about convolution neural networks (CNN)?

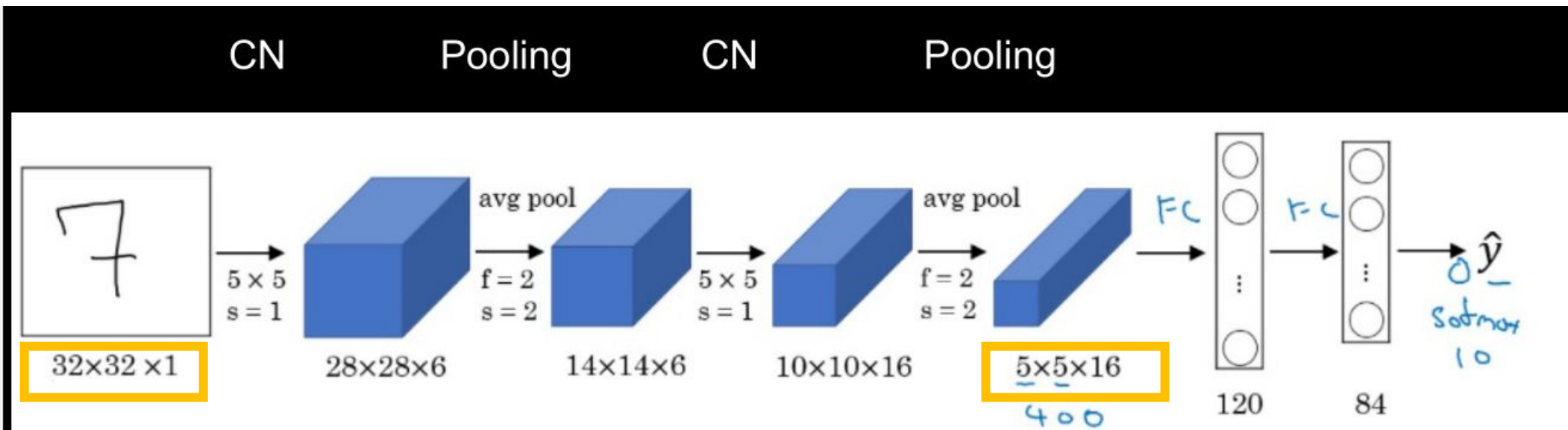
- 1. CNN have convolutional layers with filters to extract a pattern of an image**
- 2. CNN are highly sensitive to the training dataset**
3. CNN have a least one max pooling layer

# Q3

Which of the following are correct about convolution neural networks (CNN)?

## 1. CNN have convolutional layers with filters to extract a pattern of an image

The goal is to reduce the size of the image - while keeping relevant information - through multiple layers to obtain a prediction  $y$

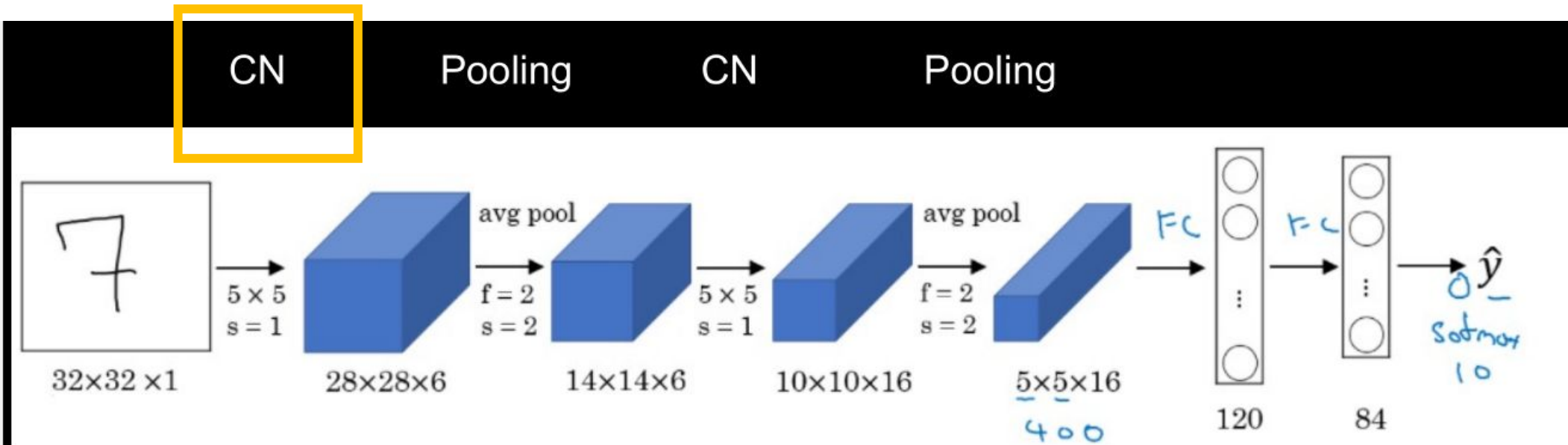


# Q3

Which of the following are correct about convolution neural networks (CNN)?

**1. CNN have convolutional layers with filters to extract a pattern of an image**

The goal is to reduce the size of the image - while keeping relevant information - through multiple layers to obtain a prediction  $y$



# Q3

Which of the following are correct about convolution neural networks (CNN)?

**1. CNN have convolutional layers with filters to extract a pattern of an image**

The goal is to reduce the size of the image - while keeping relevant information - through multiple layers to obtain a prediction  $y$

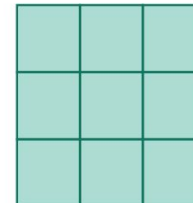
## Convolution

Image:  $x$

544	552	570	585	600	607	608	581	558	577
549	561	595	617	610	601	595	562	545	563
579	574	554	538	556	598	614	596	588	582
529	514	486	476	483	509	552	584	604	586
506	499	468	421	459	547	588	596	598	603
567	561	519	484	510	557	586	612	603	565

$$w * x$$

Convolution kernel:  $w$





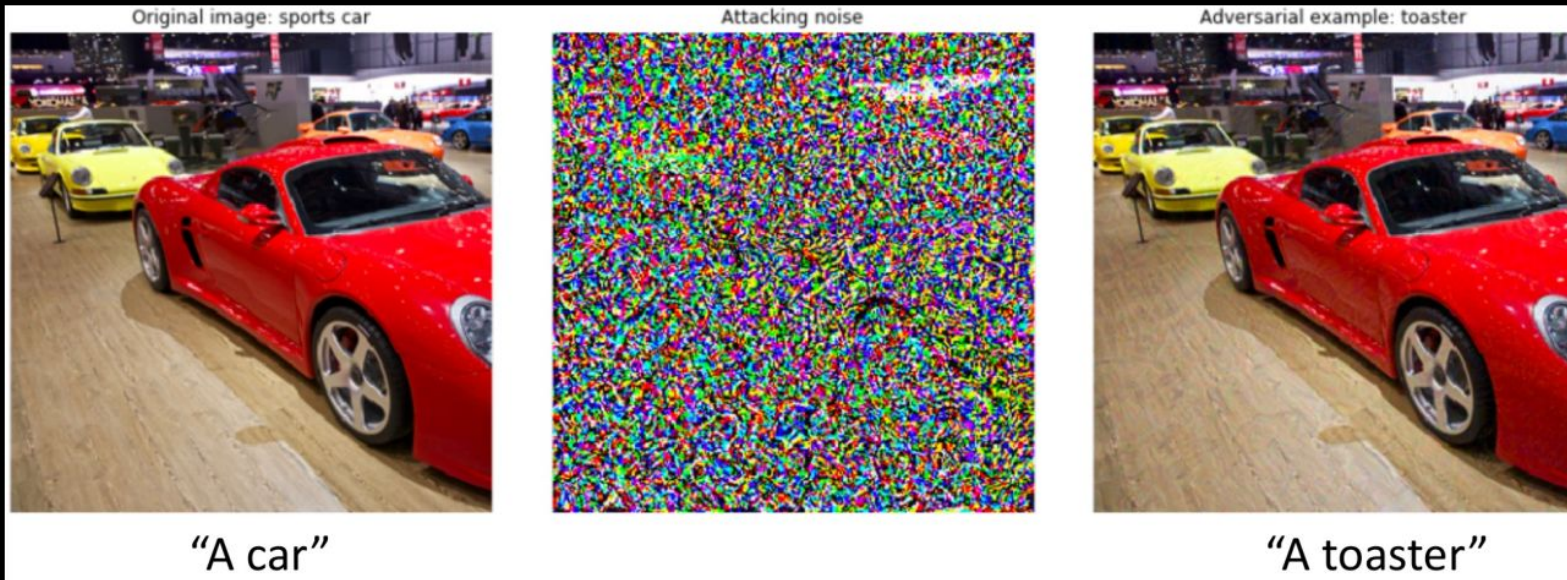
# Q3

Which of the following are correct about convolution neural networks (CNN)?

**2. CNN are highly sensitive to the training dataset**

specific perturbations in the input can generate wrong predictions

## Adversarial Attacks



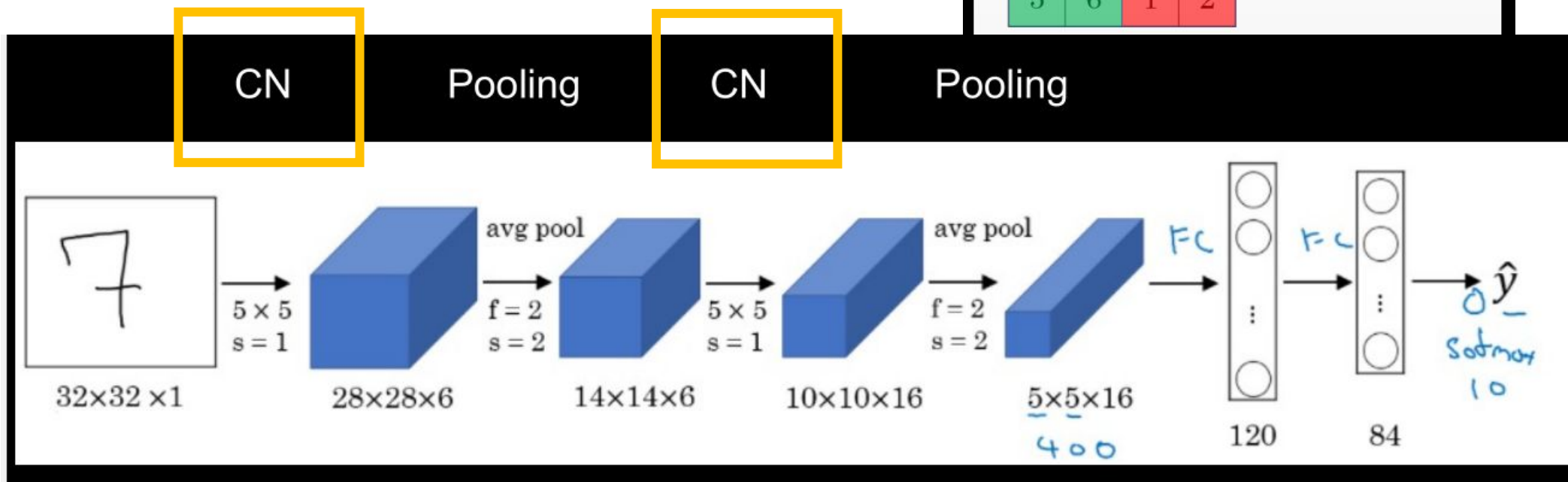
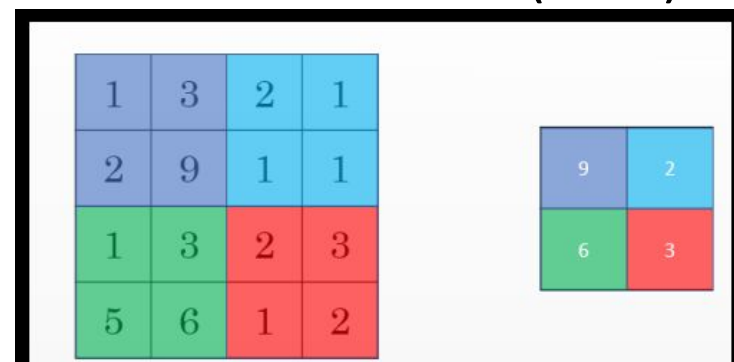


# Q3

Which of the following are correct about convolution neural networks (CNN)?

3. CNN have a least one max pooling layer

We have flexible architectural design choices in ANN



# Q4

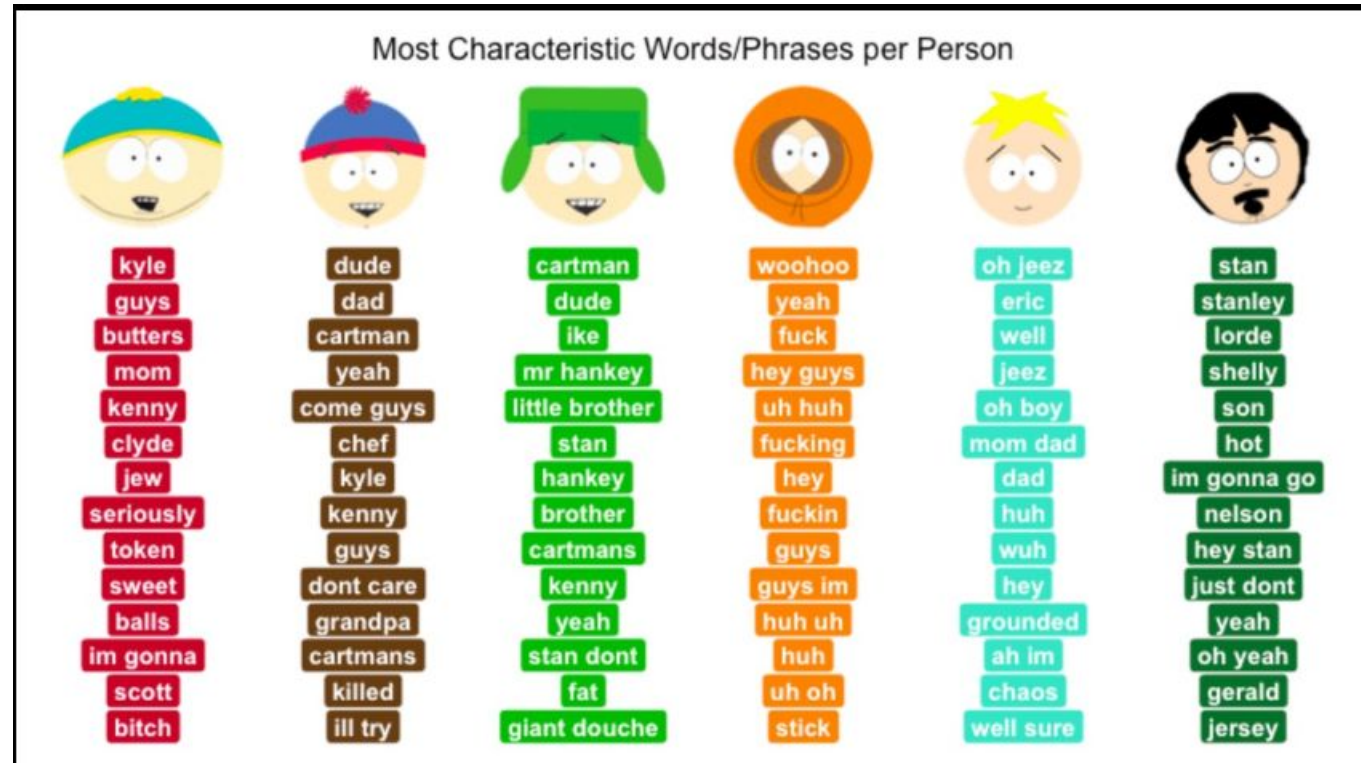
ANN are also used for text mining. Which of the following statements are correct?

- 1. In keyword extraction: given an input string, we aim to find the most relevant words in the string**
- 2. In sentiment analysis: given an input string, we aim to find a score for the most relevant words**
- 3. Topic modelling extracts relevant topics from a collection of documents**

# Q4

ANN are also used for text mining. Which of the following statements are correct?

1. In keyword extraction: given an input string, we aim to find the most relevant words in the string



# Q4

ANN are also used for text mining. Which of the following statements are correct?

**2. In sentiment analysis: given an input string, we aim to find a score for the most relevant words**

- Sentiment Analysis assigns a sentiment score to each word in a piece of text

I **love** data science, and our teachers are **awesome** +4 (Strongly positive)

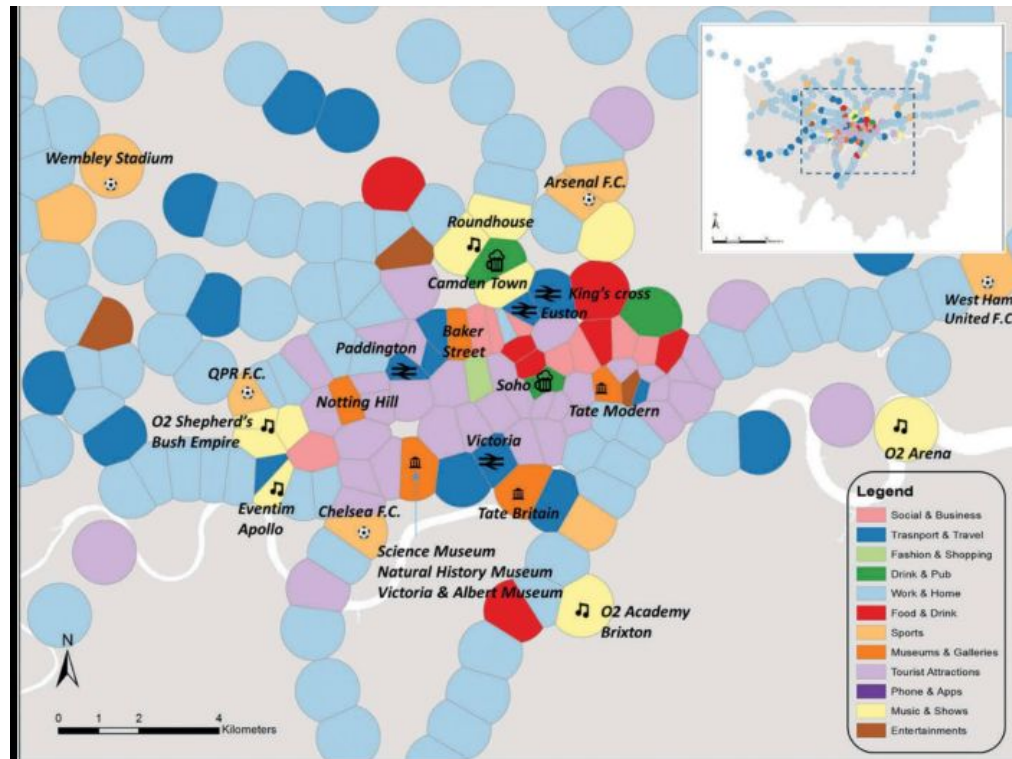
Beer is **disgusting**, why do people even **like** it? -1 (Weakly negative)

It's so **great** that my train is **late** every single day +1 (Weakly positive)

# Q4

ANN are also used for text mining. Which of the following statements are correct?

**3. Topic modelling extracts relevant topics from a collection of documents**



# Q5

Which of the following statements are true about deep learning?

1. Deep learning requires feature engineering
2. Deep learning works well with few datapoints
3. **Deep learning has neural networks with many hidden layers**
4. **ANN are prone to local minima**

# Q5

Which of the following statements are true about deep learning?

1. Deep learning requires feature engineering

We usually don't perform feature engineering in deep learning however some normalisation of the input data might be required (see also practical)

# Q5

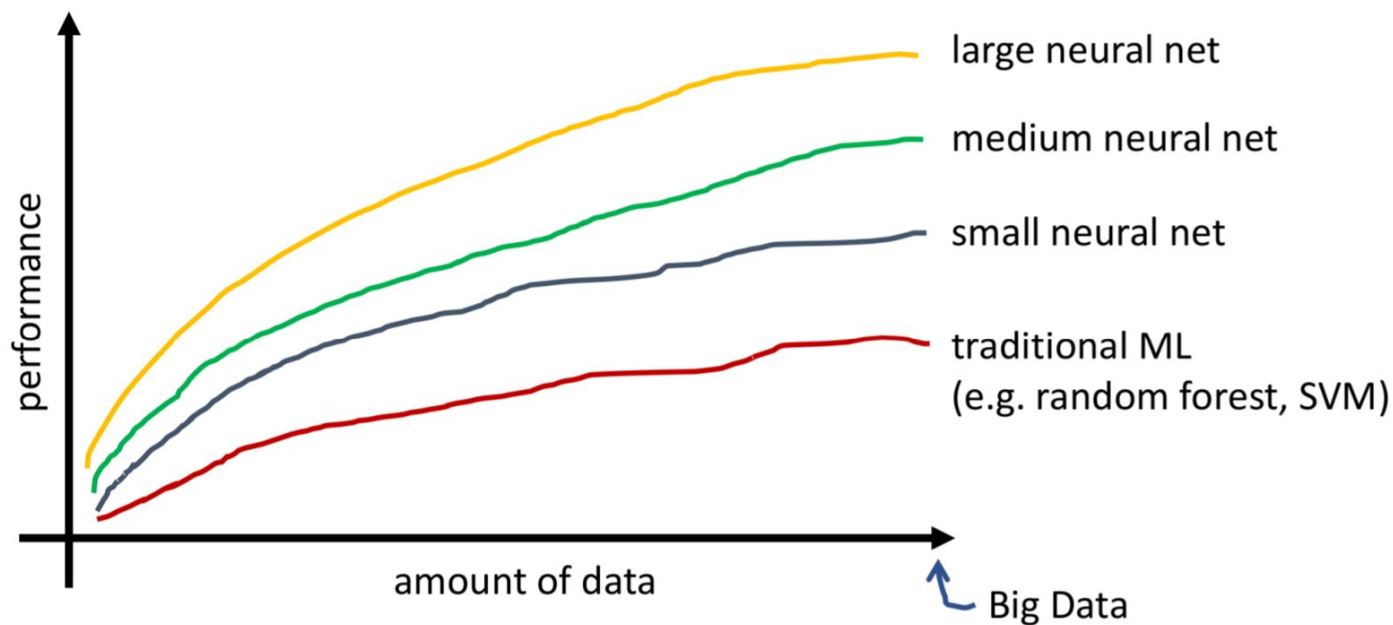
Which of the following statements are true about deep learning?

2. Deep learning works well with few datapoints

## Deep learning and data size

The performance of DL increases rapidly with the size of the data.

Due to abundance of parameters, we need big data to train the model (and avoid overfitting)





# Q5

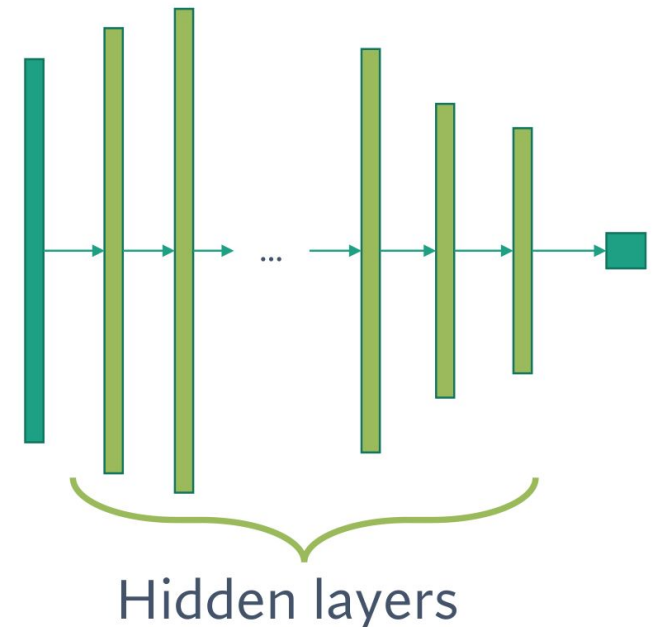
Which of the following statements are true about deep learning?

**3. Deep learning has neural networks with many hidden layers**

Deep (learning) stands for a neural network architecture with high depth, that is many hidden layers.

Increasing depth implies more parameters (weights,  $W$ ) = higher number of connections between the input (features) and the output (predictions).

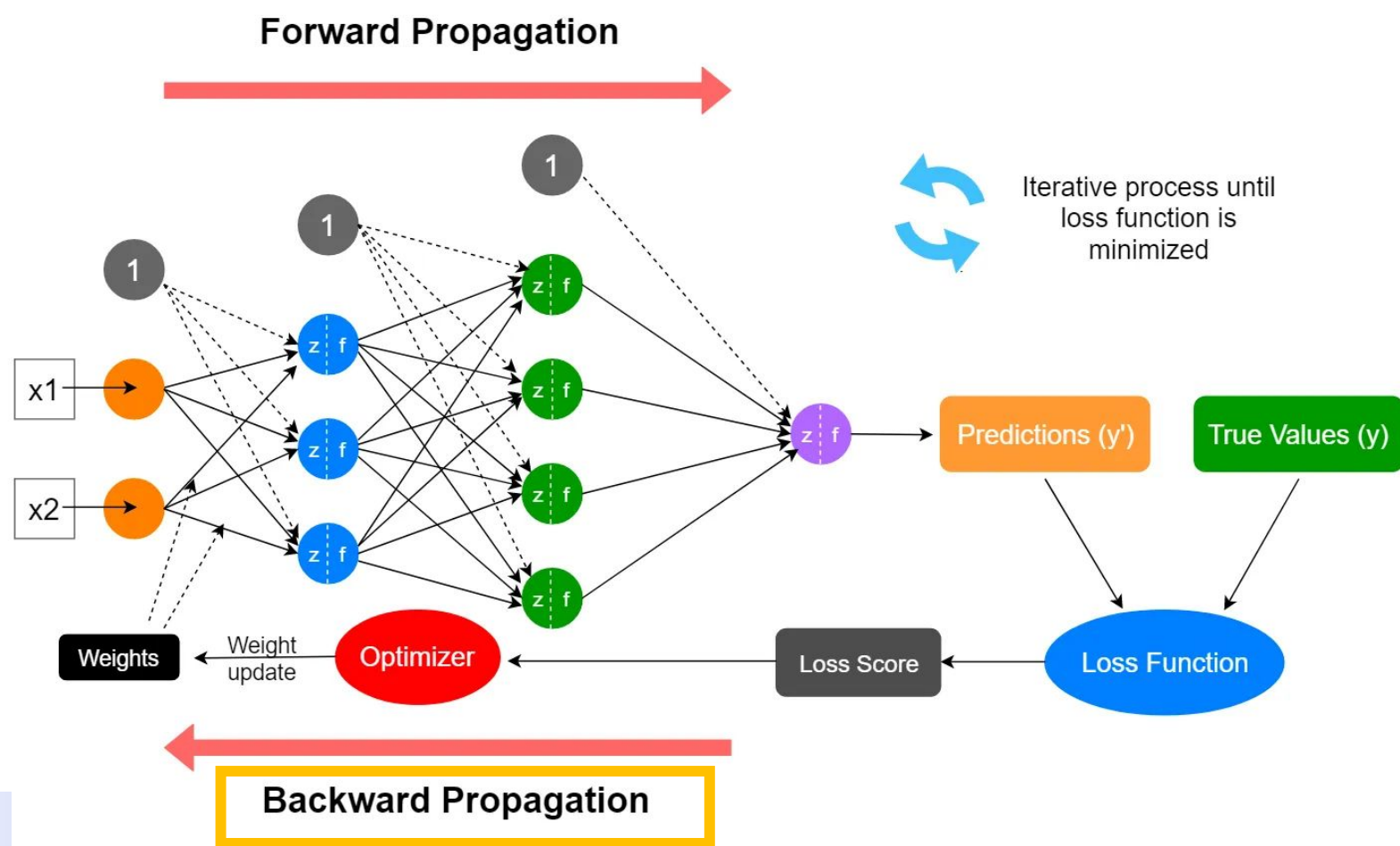
**MLP – increasing depth**



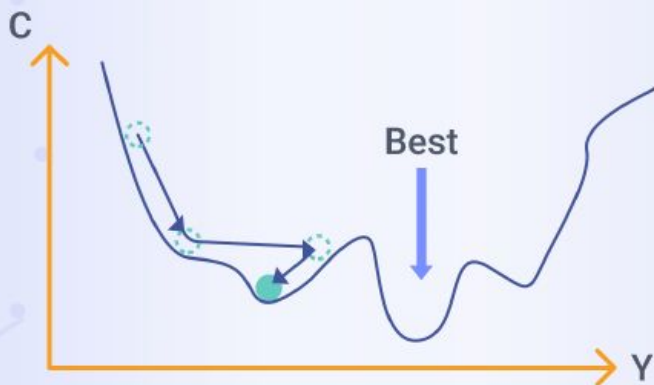
# Q5

Which of the following statements are true about deep learning?

**4. ANN are prone to local minima**



## Stochastic Gradient Descent



In the backward propagation, we use optimisation algorithms such as stochastic gradient descent which optimises the loss function BUT *are prone to local minima*: the prediction found might not be the optimal one (but DL works well in general when a big dataset is available)