[Quiz] Multi-Layer Neural Networks and Non-linear Classification

- Due 27 Apr at 23:59
- Points 10
- Questions 9
- Time limit None
- Allowed attempts 2

This quiz is no longer available as the course has been concluded.

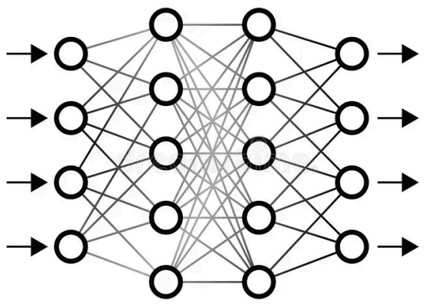
Attempt history

	Attempt	Time	Score
KEPT	Attempt 2	2 minutes	10 out of 10
LATEST	Attempt 2	2 minutes	10 out of 10
	Attempt 1	4 minutes	6 out of 10

Score for this attempt: 10 out of 10
Submitted 26 Apr at 22:48
This attempt took 2 minutes.

Question 1
1 / 1 pts

How many total layers are in the neural network schematic given below?



2
5
4

Correct!

3

Question 2
1 / 1 pts

In order for a neural network to be a theoretical "universal approximator" it must have at least 1 layer(s). Theoretically, this means a neural network can approximate any continuous function.

Answer 1:

Correct!

1
4
3
2

Question 3
1 / 1 pts

Select all that apply.

Why do we typically want to add more hidden layers to neural networks?

Theory shows us that stacking layers tends to allow for better generalization to unseen data.

Correct!

Empirical results show that stacking layers tends to allow for better generalization to unseen data.

Theory shows us that adding more hidden layers will always produces higher performing models.

Correct!

More layers tends to reduce the amount of neurons needed to represent complex data/functions.

Question 4
1 / 1 pts

If we have classification data with 10 classes, 20 features, and 100 data samples, how many neurons will the output layer need?

100
20

Correct!

10
1

Question 5
1 / 1 pts

The ReLU activation function helps to avoid the vanishing gradient problem.

Answer 1:

exploading gradient

Correct!

vanishing gradient
dying ReLU
shrinking gradient

Question 6
1 / 1 pts

Select all that apply.

We choose **NOT** to compute the derivative of the softmax directly because which of the following reasons?

We do compute the derivative of softmax directly , its derivative is 1.

Correct!

Computing the softmax directly requires addressing two different cases where the derivatives differ depending on the case, thus creating a more complex derivative.

Softmax has no derivative

Correct!

Computing the softmax together with the negative log likelihood produces a simplified and less complex derivative.

Question 7
2 / 2 pts

The main idea for generalizing the backpropagation algorithm is that we always pass on the partial derivative of the negative log likelihood (NLL) loss with respect to the _____ to the next layer in the backpropagation process.

current layer's weights $\frac{\partial NLL}{\partial \mathbf{W}^{[layer]}}$
current layer's bias $\frac{\partial NLL}{\partial \mathbf{b}^{[layer]}}$

Correct!

current layer's inputs $\frac{\partial NLL}{\partial \mathbf{A}^{[layer-1]}}$
previous layer's inputs $\frac{\partial NLL}{\partial \mathbf{A}^{[layer-2]}}$

Question 8
1 / 1 pts

Choose the best answer.

Say we are using the generalized backpropagtion algorithm for a multi-layer neural network. Before we can compute the gradient for the 1st hidden layer weights $\frac{\partial NLL}{\partial \mathbf{W}^{[1]}}$, we need to compute which of the following in **2nd** hidden layer.

- $\frac{\partial NLL}{\partial \mathbf{A}^{[2]}}$
- $\frac{\partial NLL}{\partial \mathbf{b}^{[2]}}$

Correct!

- $\frac{\partial NLL}{\partial \mathbf{A}^{[1]}}$
- $\frac{\partial NLL}{\partial \mathbf{W}^{[2]}}$

Question 9

1 / 1 pts

Select all that apply.

Which of the following activation functions suffer from the vanishing gradient problem when using multiple hidden layers.

- [ ] Softmax

Correct!

- [x] Sigmoid

Correct!

- [x] Tanh
- [ ] Identity/Linear
- [ ] ReLU

Quiz score: 10 out of 10