Module 2 Quiz

Return to Assessment List

Part 1 of 1 - / 20.0 Points

Question 1 of 10 2.0 Points

Which statement is NOT a reason why ANNs will have more impact today?

- ✔ A. Huge quantity of data available to train
- ✔ B. Increase in computing power
- ✔ C. Development of improved training algorithms
- ✔ D. Theoretical limitation of ANN turns out to be a big threat to its success
- ✔ E. A virtuous circle of funding and progress

Answer Key: D

Question 2 of 10 2.0 Points

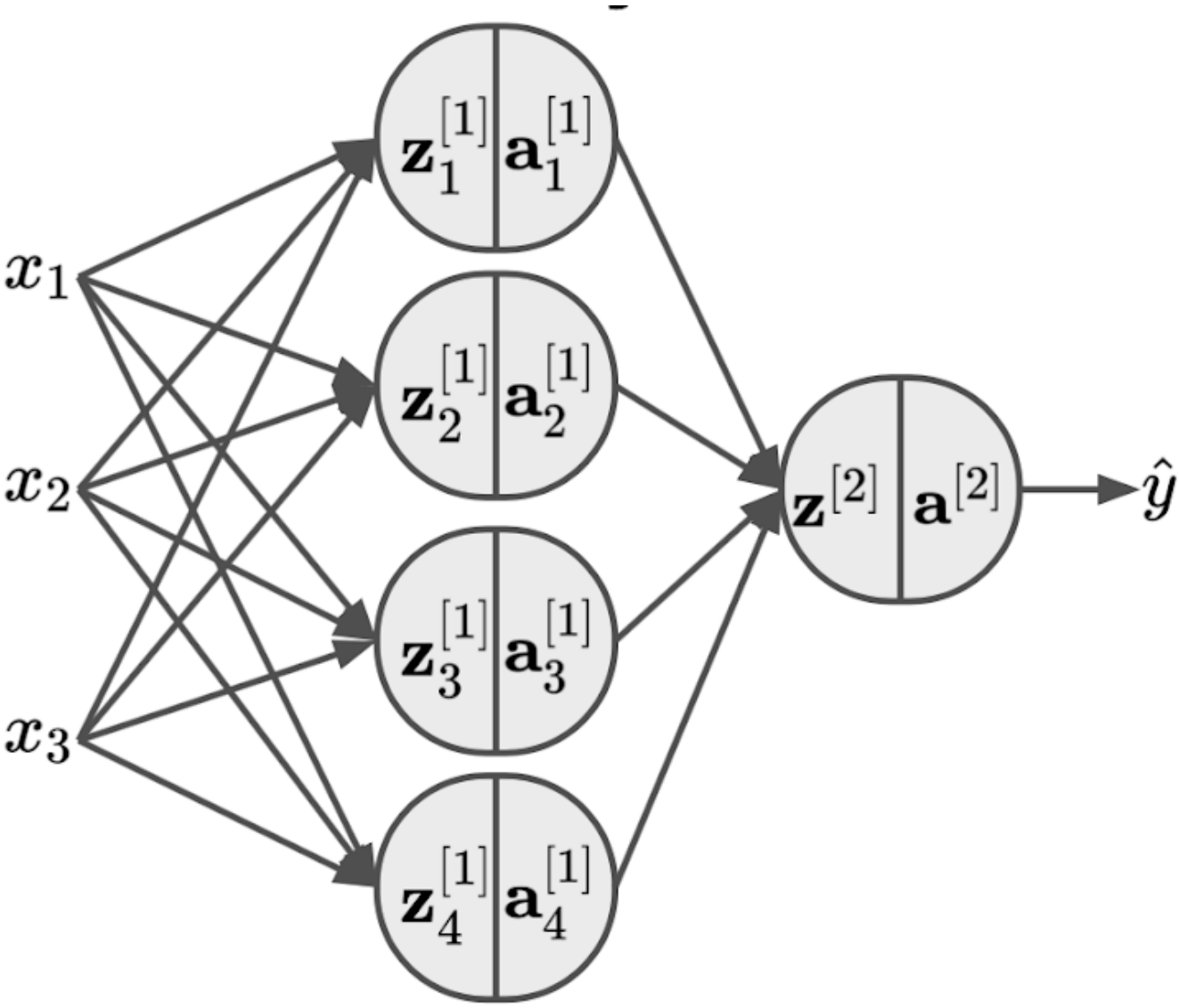
Why can the XOR classification problem not be solved using a single Perceptron?

- ✔ A. Because the Perceptron can only make predictions based on hard threshold
- ✔ B. Because the Perceptron can only make predictions based on probabilities
- ✔ C. Because the Perceptron can only make predictions on linearly separable data
- ✔ D. Because the Perceptron needs additional training data for its training
- ✔ E. All of the above

Answer Key: C

Question 3 of 10 2.0 Points

Consider the following ANN. Which are the dimensionality of W[2] in z[2] = W[2]a[1] + b[2]?



- ✔ ○ A. 4 x 3
- ✔ ○ B. 3 x 1
- ✔ ○ C. 4 x 1
- ✔ ○ D. 1 x 4
- ✔ ○ E. 1 x 1

Answer Key: D

Question 4 of 10 2.0 Points

While training the MLPs using back-propagation, the connections weights are updated in which iterative step?

- ✔ A. Forward Pass
- ✔ B. Backward Pass
- ✔ C. Gradient Descent
- ✖ D. All three steps
- ✔ E. None of the steps

Answer Key: C

Question 5 of 10 2.0 Points

Which activation function does not work well with gradient descent?

- ✔ A. ReLU Function
- ✔ B. Step Function
- ✔ C. Logit Function
- ✔ D. Tanh Function

Answer Key: B

Question 6 of 10 2.0 Points

Assume that the inputs $\mathbf{x}$ to some scalar function $\boldsymbol{f}$ are $n \times m$ matrices. What is the dimensionality of the gradient of $\boldsymbol{f}$ with respect to $\mathbf{x}$?

- ✔ A. n x m
- ✖ B. n x 1
- ✔ C. 1 x m
- ✔ D. 1 x 1

- ✔ ○ E. n x n

Answer Key: A

Question 7 of 10 2.0 Points

Why is it much more challenging to deal with multiple hyperparameters?

✔ ☐ A. There exist too many combinations of the hyperparameters

✔ ☐ B. It is too time-consuming to select the right settings of the hyperparameters

☐ C. The effect of changing a hyperparameter is linear.

✔ ☐ D. A training session can be computationally expensive (i.e. requires a lot of computational time for a trial)

Answer Key: A, B, D

Question 8 of 10 2.0 Points

What degree of polynomial do you need to reduce the training error to 0?

- ✔ ○ A. 9
- ✔ ○ B. 10
- ✔ ○ C. 20
- ✔ ○ D. 100
- ✔ ○ E. Just as same as the number of data points.

Answer Key: E

Question 9 of 10 2.0 Points

If you increase the number of iteration of training, and the validation error increase after a while. How could we fix this?

- ✔ ○ A. Stop the training when validation error is lowest
- ✔ ○ B. Have a higher learning rate

- ✔ C. Select another validation set
- ✔ D. Choose another more powerful algorithm
- ✔ E. Look for a bug in the implementation

Answer Key: A

Question 10 of 10 2.0 Points

Assume that we want to use softmax regression to predict the next word based on some features. What are some problems that might arise from a large vocabulary?

- ✔ A. It will make the probability of every word near 0.
- ✔ B. It will make the softmax function overflow.
- ✔ C. It will make the probability of the highest word near 1.
- ✔ D. It will make the probability of many words equal.

Answer Key: A