

### Question 1 10 pts

There are 3 major operations in a shallow neural network.

(A) Pass the results through an activation function.

(B) Compute the linear functions of the input data.

(C) Weight the resulting activations, take the sum, and add an offset.

What is the correct order of operations?

☐ ABC

☒ BAC

☐ BCA

### Question 2 10 pts

One of the most common activation functions is ReLU. It clips negative values to zero.

☒ True

☐ False

### Question 3 10 pts

The number of hidden units in a shallow network is a measure of the network capacity. With ReLU activation functions in  $D$  hidden units, what is the maximum number of linear regions a shallow network can have?

☐  $D$

☐  $D-1$

☒  $D+1$

### Question 4 10 pts

With enough capacity (hidden units), a shallow network can describe any continuous 1D function defined on a compact subset of the real line to arbitrary precision.

☒ True

☐ False

### Question 5 10 pts

We can extend the deep network construction to more than two hidden layers; modern networks might have more than a hundred layers with thousands of hidden units at each layer. Find the terms that match the concepts.

The number of hidden units.

The number of hidden layers.

The total number of hidden units.

### Question 6 10 pts

Which of the following is equivalent to the maximum likelihood criterion?

☐ Maximizing log-likelihood.

☒ Minimizing the negative log-likelihood.

☐ Minimizing the cross-entropy loss.

**Question 7****10 pts**

Select all the correct statements below.

- ☒ We often use the word "layers" to refer to the structure of neural networks, e.g., input layer, hidden layer, and output layer.
- ☒ The hidden units themselves are sometimes referred to as neurons.
- ☒ Any neural network with at least one hidden layer is also called a multi-layer perceptron, or MLP for short.
- ☒ Networks with one hidden layer are sometimes referred to as shallow neural networks.

**Question 8****10 pts**

If we pass the output of one shallow neural network into a second shallow network. This is a special case of a deep network with two hidden layers.

- ☒ True
- ☐ False

**Question 9****10 pts**

When building a neural network for K-class classification, we often pass the K outputs of the network through the softmax function.

- ☒ True
- ☐ False

**Question 10****10 pts**

Select the assumptions we have in the maximum likelihood method.

- ☒ Assume that the data are identically distributed (the form of the probability distribution over the outputs  $y_i$  is the same for each data point).
- ☒ Assume that the conditional distributions  $\Pr(y_i|x_i)$  of the output given the input are independent.