1. Each input in the MNIST classification problem has 28 x 28 pixels and each pixel can take values in the integers 0 through 255 (inclusive). Let's call the space of all such inputs the MNIST total input space. What is the size of the MNIST total input space?

- (●) 256^784 where "^" denotes exponentiation
- ( ) 60000
- ( ) 10000
- ( ) 256*784

2.

Each input in the MNIST classification problem has 28 x 28 pixels and each pixel can take values in the integers 0 through 255 (inclusive). Let's call the space of all such inputs the MNIST total input space.

Let w be a 28 x 28 NumPy 2-D array of non-zero weights (weights are floating point numbers and can be positive or negative). For this w, which of the following expressions correctly computes **the input** in the MNIST total image space **that has the MAXIMUM inner product** (aka dot product) **with w**? For computing the inner product, think of both the 28x28 input and the 28x28 array of weights as 784 dimensional vectors.

Hint: think about what pixel value you should set in your 28 x 28 input for pixels where w has a positive weight. And what about pixels where w has a negative weight?

- (●) 255*(w > 0)
- ( ) There is no simple expression for the inner product maximizing input. It can only be computed numerically via gradient descent.
- ( ) w > 0
- ( ) w

3.

This model building code was written to solve a classification problem with K classes. What was the value of K?

```
model = keras.Sequential([
    layers.Dense(512, activation="relu"),
    layers.Dense(128, activation="relu"),
    layers.Dense(64, activation="softmax")
])
```

- (●) 64
- ( ) 128
- ( ) 512
- ( ) value of K cannot be determined from the code given above

4.

What is the core building block of neural networks?

- (●) layer
- ( ) loss
- ( ) activation function
- ( ) optimizer

5.

What does the term "overfitting" refer to?

- (●) Machine learning models tend to perform better on training data than on test data they have never seen
- ( ) Machine learning models need a lot of computation to successfully train
- ( ) Machine learning models have a large number of training parameters in them
- ( ) Machine learning models for one data type (e.g., images) do not do well on another data type (e.g., text)

6.

Suppose I have a vocabulary of size 10 with the word index given below. What will be the multi-hot encoding of the sentence: "we love deep learning"?

you, 0
we, 1
us, 2
love, 3
like, 4
deep, 5
shallow, 6
learning, 7
teaching, 8
machine, 9

- ( ) the list [1, 3, 5, 7]
- (●) the vector (0, 1, 0, 1, 0, 1, 0, 1, 0, 0)
- ( ) the list [1, 3, 5, 7, 0, 0, 0, 0, 0, 0]
- ( ) the vector (1, 0, 1, 0, 1, 0, 1, 0, 0, 0)

7

Suppose your inputs have 17 features and you have a classification problem with 10 categories. A softmax regression model in Keras for this task would require just a single layer. What will be the code to create that single layer?

- (●) layers.Dense(10, activation = "softmax")
- ( ) layers.Dense(10)
- ( ) layers.Dense(17)
- ( ) layers.Dense(10, activation = "relu")

8.

Which of the following correctly describes the ReLU activation function?

- ( ) |x|
- ( ) min{0, x}
- (●) (x + |x|) / 2
- ( ) None of the above

9.

What does the "M" in "MNIST" stand for?

- (●) modified
- ( ) multiclass
- ( ) machine
- ( ) multiple

10.

Suppose your inputs have 13 features and you have a real valued label/target (like house price) that you want to predict. A linear regression model in Keras for this task would require just a single layer. What will be the code to create that single layer?

- (●) `layers.Dense(1)`
- ( ) `layers.Dense(1, activation = "relu")`
- ( ) `layers.Dense(1, activation = "sigmoid")`
- ( ) `layers.Dense(13)`