# Test 1 Suggested Questions

``Describe the method for gradient descent used to determine weights

that minimize error?''

1. What is linear regression and how is it applied in machine

learning.

2. Linear Unit Function VS Sigmoid Function.

## Fill in the blank (6)

1. Although there are formulas we can solve to find the optimal weights for \_\_\_\_\_\_\_\_\_ and \_\_\_\_\_\_\_\_\_ models, we must use \_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_ to find the optimal weights for \_\_\_\_\_\_\_\_\_.
2. If the machine is complex (has thousands of weights), it will likely be \_\_\_\_\_\_\_\_\_fitting. If the machine is not complex enough to model the data, it will likely be \_\_\_\_\_\_\_\_\_fitting.

## Short Answer:

Steps of Supervised Machine Learning to Process an Image.

Fill in the blank and put the statements in the right order.

1. Machine’s answers are plotted against the \_\_\_\_\_\_\_\_\_\_\_\_ answers Yd.
2. Y is determined by choosing the most \_\_\_\_\_\_\_\_\_\_\_\_ solution. Y is the machine’s best \_\_\_\_\_\_\_\_\_\_\_\_ to the training data.
3. Machine calculates z by \_\_\_\_\_\_\_\_\_\_\_\_ each \_\_\_\_\_\_\_\_\_\_\_\_ multiplied by each input vector element
4. The machine computes the probability of each possible solution occurring.
5. <fill in>
6. Fix/set the \_\_\_\_\_\_\_\_\_\_\_\_
7. Feed the \_\_\_\_\_\_\_\_\_\_\_\_ the new \_\_\_\_\_\_\_\_\_\_\_\_ without any answers.
8. Take the difference between the \_\_\_\_\_\_\_\_\_\_\_\_ answer and the \_\_\_\_\_\_\_\_\_\_\_\_’s answer, and \_\_\_\_\_\_\_\_\_\_\_\_ it.
9. \_\_\_\_\_\_\_\_\_\_\_\_ (Sigmoid) function is applied on z (or more recently, the \_\_\_\_\_\_\_\_\_\_\_\_(rectifier) function) to convert data to a curve continuous between (\_\_\_\_\_\_\_\_\_\_\_\_ , \_\_\_\_\_\_\_\_\_\_\_\_) .
10. Sum the \_\_\_\_\_\_\_\_\_\_\_\_ of all of the training data to find the \_\_\_\_\_\_\_\_\_\_\_\_ function.
11. Machine chooses arbitrary \_\_\_\_\_\_\_\_\_\_\_\_ for each pixel’s image, signifying the pixel’s importance.
12. Apply the minimization function ArgMin to find the \_\_\_\_\_\_\_\_\_\_\_\_ that minimize error
13. Give new input to the machine. This is called \_\_\_\_\_\_\_\_\_\_\_\_ data.

2. Explain why the error is squared in the error function (2):

E(w) = ∑d = 1...D (yd - Mw(Xd))2

3. Label and draw on the graph the error of data point Xd

Error of one point = yd - Mw(Xd)

xd

Input (x)

Output (y)

4. What does this equation mean?:

m = (epsilon)D

5. What two values does a weight w depend on?

6. What is the benefits of applying the sigmoid function to z (compared to the threshold function)

7. What is the benefits of applying the sigmoid function to z (compared to the rectifier function)

8. Given the following set of Md,i What is the photo I likely of?

Md,1 = 0.02 cat

Md,2 = 0.60 face

Md,3 = 0.35 dog

Md,4 = 0.03 car

## True/False

If false, explain why.

1. Error can be calculated for general (new) data
2. We choose a weight based on: w = ArgMinw E(w) + Complexity(Mw) Because we want the machine to minimize the error and be as complex as possible.