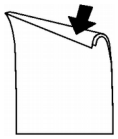
CSCE 5063 Spring 2018 Midterm 1

1. (5 points) I plan to return these tests face-down, so please write your name on the back, near the top.



2. (5 points) Circle the statement that is more consistent with the teachings of this class:

- When optimizing more weights, training time becomes a bigger problem than local optima.
- When optimizing more weights, local optima becomes a bigger problem than training time.

3. (5 points) Circle the statement that is more consistent with the teachings of this class:

- When tuning meta-parameters (such as learning rate, momentum term, number of layers, size of layers, etc.), intuition trumps cross-validation.
- When tuning meta-parameters, cross-validation trumps intuition.

4. (5 points) Circle the statement that is more consistent with the teachings of this class:

- Empirical gradient descent is more efficient than backpropagation.
- Backpropagation is more efficient than empirical gradient descent.

5. (5 points) Circle the statement that is more consistent with the teachings of this class:

- The initial weights have little effect on neural networks when training with gradient descent.
- Proper weight initialization is important making gradient descent work.

6. (5 points) Circle the statement that is more consistent with the teachings of this class:

(continued in the next column...)

- A neural network with two linear layers and one non-linear layer can fit to pretty-much any dataset with arbitrarily low error.
- A neural network can fit to any data with low error, but only if you find the right activation function and the right number of layers.

7. (5 points) Circle the statement that is more consistent with the teachings of this class:

- With linear models, all reasonable optimization methods should arrive at approximately the same set of weights.
- With linear models, OLS will find a better optimum than most other optimization methods.

8. (5 points) Circle the statement that is more consistent with the teachings of this class:

- The ultimate goal in machine learning is usually to generalize.
- The ultimate goal in machine learning is usually to fit to the training data as closely as possible.

9. (5 points) If you plot accuracy over time (as we did in assignments 3), and the accuracy jumps up and down excessively, a good way to smooth it out might be:

- add some momentum and decrease the learning rate.
- set the momentum to zero and increase the number of folds.

10. (5 points) Circle the statement that is more consistent with the teachings of this class:

- To derive gradient descent, we compute the derivative of the error metric with respect to the derivative of the weights.
- To derive gradient descent, we compute the derivative of the model with respect to the derivative of the weights.

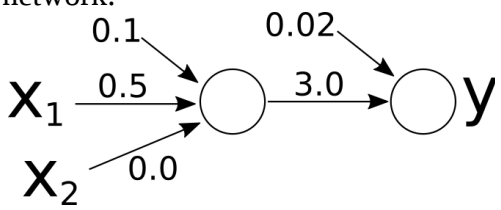
11. (5 points) Please briefly identify 3 ways to limit over-fit with a neural network:

1.

2.

3.

12. (5 points) Consider this simple two-layer neural network:



The two numbers at the top of this Figure are bias terms ("b"). The three other numbers are linear weights ("M").

The activation function used in both units is $f(x) = x^2$. The derivative of this activation function is $f'(x) = 2x$.

If you feed the feature vector [x_1=0.2, x_2=0.7] into this neural network, what value will be predicted for y? (Please double-check your answer to this question, because the next few depend on it.)

13. (5 points) Suppose the training pattern [x_1=0.2,x_2=0.7] → [y=0.24] is presented to this neural network for training by stochastic gradient descent. (Assume the objective function is squared error, just like in your programming assignments. The learning rate is $\eta = 0.1$. The momentum term is $\mu = 0.0$.)

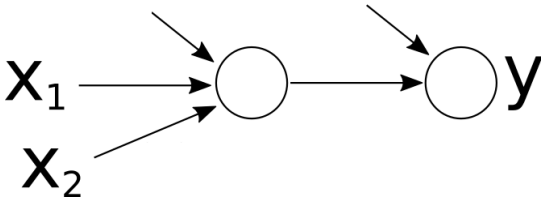
What "blame term" will backpropagation assign to the output of the output unit?

14. (5 points) What "blame term" will backpropagation assign to the net input of the output unit?

15. (5 points) What blame term will backpropagation assign to the output of the hidden unit?

16. (5 points) What blame term will backpropagation assign to the net input of the hidden unit?

17. (5 points) What will the updated weights be? (The biases are weights too, so your answer should report 5 values. Put your final answer on the figure below.)



18. (15 points) Consider the following table of data. (Admittedly, this data is not big enough to represent anything serious.)

Features		Labels
Cranium circumference	Favorite animal	A.C.T Score
21.76	Elephant	26
34.62	Chigger	30
24.01	Elephant	24

(This question is continued in next column…)

If you perform one repetition of 3-fold cross-validation with the baseline learning algorithm using this data, what will be the resulting sum-squared error?

(At training time, the baseline algorithm ignores the training features and computes the mean of the labels. At prediction time, it ignores the features and predicts the mean label value.)