Midterm 2 Study Guide

Linear Regression Models

1. What effect does the bias term have on the model's predictions?

2. Explain how to decide whether sigmoid or softmax is used in a regression model.

3. How many weights are there in a linear regression model that uses 3 features to represent poems, and predicts from a set of 5 possible labels?

4. Explain what each of the terms in the equation for cross-entropy loss represents.

Training Supervised Machine Learning Classifiers

5. What is an *epoch* of training?

6. What would a learning rate of zero mean?

7. ReLu is a non-linear activation function that returns the z-score if it is greater than 0, and 0 otherwise. Can we use ReLu as the activation function for the final output unit in a classification model? Why or why not?

8. What does *fully-connected* mean in the context of neural networks?

9. A single unit neural network is used to classify comments as positive or negative. The three input features are word count, count of positive words, and count of negative words. After seeing Comment A with the feature representation [0, 0, 1], the model assigns 80% probability to the negative class.

    A. If the true label is negative, what is the cross-entropy loss?

    B. Would you expect the weight on negative words to go up or down after one step of gradient descent on this input?

10. Explain the difference between the term *parameter* and the term *hyperparameter.*

Feature Representation

11. How are grayscale images represented as input features for a supervised machine learning classifier?

12. How are color images represented as input features for a supervised machine learning classifier?

13. What are *embeddings* and why are they useful?

Simple Neural Networks

14. Write the equation that computes the output from a single neural network unit that uses sigmoid as a non-linear activation function.

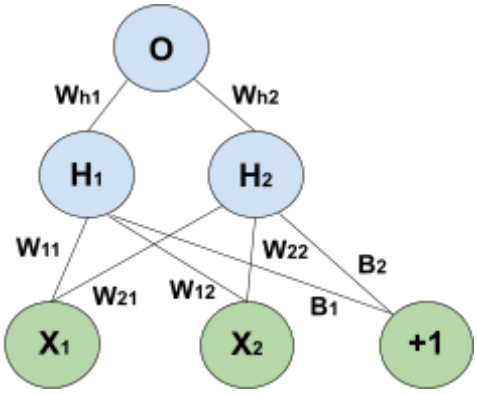
15. Can a single perceptron unit be used to solve AND? Why or why not?

16. Can a single perceptron unit be used to solve XOR? Why or why not?

17. Explain the notational differences between the equation for regression and the equation for the output of a single neural network unit.

Multi-Layer Neural Networks

Consider the network below:



18. Which layers in this network contain a bias term?

19. How many trainable weights are there in this network?

20. Could this network be used to solve AND? Why or why not?

21. Could this network be used to solve XOR? Why or why not?

22. Assume that the hidden layer uses ReLU as a non-linear activation function. Write out the equation that each unit in the hidden layer uses to compute its output.

23. For binary classification, what non-linear activation function should the output unit use? Write out the equation that computes its output.

24. Here is a set of weights for this network. Compute the score for input (1,1) if there is no non-linear activation function applied in the hidden layer, and the output layer uses the simple perceptron activation function: f(z) returns 1 if z is greater than 0 and 0 otherwise.

$W_{11} = 1$	$B_1 = -1$	$W_{21} = -0.5$	$W_{h1} = 0.5$
$W_{12} = 0.5$	$B_2 = 0$	$W_{22} = 2$	$W_{h2} = -1$

25. Draw a multi-class classification neural network that takes a vector of 4 features as input and has two hidden layers. The first hidden layer has three units and the second hidden layer has two units. Show and label all weights.

26. You are training a binary classification model with a single unit hidden layer and an output layer. Give the equations that each unit uses to compute its output.