Quiz

Name:

You can use 5 pages of notes for this quiz.

To find out what questions to answer, divide the last 2 digits of your UTEP ID by 6.
* If the remainder is 0, answer questions 6, 12, 18, and 24
* If the remainder is 1, answer questions 1, 7, 12, and 19
* If the remainder is 2, answer questions 2, 8, 13, and 20
* If the remainder is 3, answer questions 3, 9, 14, and 21
* If the remainder is 4, answer questions 4, 10, 15, and 22
* If the remainder is 5, answer questions 5, 11, 16, and 23

1. If after training, we get the values of 5 parameters each of which is described by a 16-bit number, how many bits do we need to store the information about all possible combinations of these parameters?

2. How many possible 8-bit binary sequences can we have?

3. In general, how many different models can you have if, to describe each model, you use 2 variables with 4 bits each?

4. If we have 4 neurons, how many possible permutations can we have that do not change the model?

5. How many models can be represent by a traditional neural network of K neurons that uses N bits to store all the weights? Explain your answer.

6. If we use 4 neurons each of which uses 2 bits, how many possible models can we represent? How many models could we represent in general if we use the same number of bits?

7. Explain why, to get a better accuracy, we need to go from the traditional neural network with many neurons in a single layer to a multi-layer structure with a small number of neurons in each layer.

8-9. Explain, in detail, why for deep neural networks, the activation function should be piece-wise linear.

10. Why should we have c+ different from c- in the activation function?

11. What activation function is actually used in deep learning? Why?

12. Prove that every piece-wise linear function can be represented as a linear combination of the rectified linear function and the identity function.

13. Represent the following function as a linear combination of the rectified linear function and the identity function:
* for $z>0$, we have $s(z)= 2z$;
* for $z<0$, we have $s(z)=z$.

14-15. Why cannot we require shift-invariance instead of scale-invariance?

16. Why do we need pooling?

17-18. Which pooling operations work most successfully in deep learning? Provide a theoretical explanation for these operations.

19. Suppose that instead of the three operations, we use the operation $p(x,y) = a * \max(x,y) + (1 - a) * \min(x,y)$ with $a = 0.7$.

Provide an example that for this operation, the result of pooling four numbers may depend on how we divide the original four numbers into pairs.

20. Explain what researchers mean when they say that deep learning is too sensitive. Explain why this is a big problem in practical applications.

21. Explain why deep neural networks are very vulnerable to the noise in the original data, much more vulnerable than the traditional neural networks.

22. If neurons on each layer amplify the noise by a factor of 2, how much will the noise be amplified after passing a 6-layer neural network?

23. How to deal with sensitivity in deep learning?

24. What is the main remaining problem of big learning? Why is interpretability a problem? Explain.