

六方云算法工程师笔试题

考试说明：答题时超过 **30秒** 没有新操作就强制交卷，最多允许切屏 **1次** 超出该次数会取消继续作答权限。请同学们注意笔试要求，公正做题。

* 基本信息：

姓名：

学校及专业：

电话：

邮箱：

一、单选题

* 1. Suppose you have built a neural network. You decide to initialize the weights and biases to be zero. Which of the following statements are True? (Check all that apply)

- ☐ Each neuron in the first hidden layer will perform the same computation. So even after multiple iterations of gradient descent each neuron in the layer will be computing the same thing as other neurons.
- ☐ Each neuron in the first hidden layer will perform the same computation in the first iteration. But after one iteration of gradient descent they will learn to compute different things because we have “broken symmetry”.
- ☐ Each neuron in the first hidden layer will compute the same thing, but neurons in different layers will compute different things, thus we have accomplished “symmetry breaking” as described in lecture.
- ☐ The first hidden layer’s neurons will perform different computations from each other even in the first iteration; their parameters will thus keep evolving in their own way.

* 2. What is t-SNE?

- ☐ A non-linear dimensionality reduction technique.
- ☐ A linear transformation that allows us to solve analogies on word vectors.
- ☐ A supervised learning algorithm for learning word embeddings.
- ☐ An open-source sequence modeling library.

* 3. You have an input volume that is 15x15x8, and pad it using “pad=2.” What is the dimension of the resulting volume (after padding)?

- ☐ 17x17x10
- ☐ 19x19x8
- ☐ 19x19x12
- ☐ 17x17x8

* 4. You have trained word embeddings using a text dataset of m1 words. You are considering using these word embeddings for a language task, for which you have a separate labeled dataset of m2 words. Keeping in mind that using word embeddings is a form of transfer learning, under which of these circumstance would you expect the word embeddings to be helpful?

- ☐ m1 << m2
- ☐ m2 << m1

* 5. Suppose you are training a LSTM. You have a 10000 word vocabulary, and are using an LSTM with 100-dimensional activations a. What is the dimension of Γu at each time step?

- ☐ 1
- ☐ 100
- ☐ 300
- ☐ 10000

* 6. You are building a binary classifier for recognizing cucumbers ($y=1$) vs. watermelons ($y=0$). Which one of these activation functions would you recommend using for the output layer?

- ☐ ReLU
- ☐ Leaky ReLU
- ☐ sigmoid
- ☐ tanh

二、多选题

* 7. Increasing the parameter keep_prob from (say) 0.5 to 0.6 will likely cause the following: (Check the two that apply) 【多选题】

- ☐ Increasing the regularization effect.
- ☐ Reducing the regularization effect.
- ☒ Causing the neural network to end up with a higher training set error.
- ☐ Causing the neural network to end up with a lower training set error.

* 8. Which of the following statements are true? (Check all that apply) 【多选题】

- ☐ If the training and test errors are about the same, adding more features will not help improve the results.
- ☐ If a learning algorithm is suffering from high variance, adding more training examples is likely to improve the model.
- ☒ If a learning algorithm is suffering from high bias, only adding more training examples may not improve the test error significantly.
- ☐ A model with more parameters is more likely to over-fitting and typically has high variance.

* 9. Which of these equations do you think should hold for a good word embedding? (Check the two that apply) 【多选题】

- ☐ $e_{boy} - e_{girl} \approx e_{brother} - e_{sister}$
- ☐ A
- ☒ $e_{boy} - e_{girl} \approx e_{sister} - e_{brother}$
- ☒ B
- ☐ $e_{boy} - e_{brother} \approx e_{girl} - e_{sister}$
- ☐ C
- ☐ $e_{boy} - e_{brother} \approx e_{sister} - e_{girl}$
- ☐ D

* 10. Why do we normalize the input x to Neural Network? 【多选题】

- ☐ It makes it easier to visualize the data.
- ☒ It makes the parameter initialization faster.
- ☐ It makes the cost function faster to optimize.
- ☐ Normalization is another word form regularization – it helps to reduce variance.

三、判断题

* 11. L2范数会使权值稀疏。

- ☐ 对
- ☐ 错

* 12. 当你增加正则化超参数lambda时权重会变得更

- ☐ 对
- ☐ 错

* 13. $\alpha = e^t * \alpha_0$ 不是一个好的学习率衰减方法。

- ☐ 对
- ☐ 错

* 14. 如果在大量的超参数中搜索最佳的参数值，那么应该尝试在网格中搜索而不是使用随机值，以便更系统的搜索，而不是依靠运气。

- ☐ 对
- ☐ 错

* 15. Adam算法不能用于mini-batch梯度计算。

- ☐ 对
- ☐ 错

* 16. 每个超参数如果设置得不好，都会对训练产生巨大的负面影响，因此所有的超参数都同等重要

- ☐ 对
- ☐ 错

* 17. 神经网络的深层通常比浅层计算输入的更复杂的特性。

- ☐ 对
- ☐ 错

* 18. L0范数是指向量中非0的元素的个数。

- ☐ 对
- ☐ 错

* 19. 在不同的mini-batch下，不需要显式地进行循环，就可以实现mini-batch梯度下降，从而使算法同时处理所有的数据（矢量化）

- ☐ 对
- ☐ 错

* 20. 向量化允许我们在L层神经网络中计算前向传播，而不需要在层($l = 1, 2, \dots, L$)上显式的使用for-loop（或任何其他显式迭代循环）。

- ☐ 对
- ☐ 错

四、问答题

* 21. 如果你的模型在训练集上表现很好，在新的数据上表现很差，发生了什么？你能说出可能的解决办法吗？

* 22. 你能总结一下“动态规划”问题的解体思路吗？

* 23. 使用gradient descent来训练模型的话经常要在数据预处理步骤进行数据标准化，为什么要进行数据标准化？

* 24. 归一化，标准化和白化，在原理上有哪些区别？

五、编程题

* 25. 给定一个按照升序排列的整数数组 **nums**，和一个目标值 **target**。找出给定目标值在数组中的开始位置和结束位置。
你的算法时间复杂度必须是 $O(\log n)$ 级别。
如果数组中不存在目标值，返回 **[-1, -1]**。
示例 1：
输入: **nums = [5,7,7,8,8,10]**, **target = 8**
输出: **[3,4]**
示例 2：
输入: **nums = [5,7,7,8,8,10]**, **target = 6**
输出: **[-1,-1]**

提交

举报