

机器学习与深度学习常见面试题（上）



一年一度的校园招聘已经开始了,为了帮助参加校园招聘、社招的同学更好的准备面试, SIGAI 整理出了一些常见的机器学习、深度学习面试题。理解它们, 对你通过技术面试非常有帮助, 当然, 我们不能只限于会做这些题目, 最终的目标是真正理解机器学习与深度学习的原理、应用。

1. 比较 Boosting 和 Bagging 的异同

二者都是集成学习算法, 都是将多个弱学习器组合成强学习器的方法。

Bagging: 从原始数据集中每一轮有放回地抽取训练集, 训练得到 k 个弱学习器, 将这 k 个弱学习器以投票的方式得到最终的分类结果。

Boosting: 每一轮根据上一轮的分类结果动态调整每个样本在分类器中的权重, 训练得到 k 个弱分类器, 他们都有各自的权重, 通过加权组合的方式得到最终的分类结果。

2. 无监督学习中存在过拟合吗?

存在。我们可以使用无监督学习的某些指标或人为地去评估模型性能, 以此来判断是否过拟合。

3. 什么是 k 折交叉验证?

将原始数据集划分为 k 个子集, 将其中一个子集作为验证集, 其余 $k-1$ 个子集作为训练集, 如此训练和验证一轮称为一次交叉验证。交叉验证重复 k 次, 每个子集都做一次验证集, 得到 k 个模型, 加权平均 k 个模型的结果作为评估整体模型的依据。

4. 关于 k 折交叉验证, 需要注意什么?

k 越大, 不一定效果越好, 而且越大的 k 会加大训练时间; 在选择 k 时, 需要考虑最小化数据集之间的方差, 比如对于 2 分类任务, 采用 2 折交叉验证, 即将原始数据集对半分, 若此时训练集中都是 A 类别, 验证集中都是 B 类别, 则交叉验证效果会非常差。

5. 对于一个二分类问题, 我们定义超过阈值 t 的判定为正例, 否则判定为负例。现在若将 t 增大, 则准确率和召回率会如何变化?

准确率 = $TP / (TP + FP)$, 召回率 = $TP / (TP + FN)$, 其中 TP 表示将正例正确分类为正例的数量, FP 表示将负例错误分类为正例的数量, FN 表示将正例错误分类为负例的数量。准确率可以理解为在所有分类为正例的样品中, 分类正确的样本所占比例; 召回率可以理解为在所有原始数据集中的正例样品中, 正确挑出的正例样本的比例。

因此若增大阈值 t ，更多不确定（分类概率较小）的样本将会被分为负例，剩余确定（分类概率较大）的样本所占比例将会增大（或不变），即正确率会增大（或不变）；若增大阈值 t ，则可能将部分不确定（分类概率较小）的正例样品误分类为负例，即召回率会减小（或不变）。

6. 以下关于神经网络的说法中，正确的是（ ）？

- A. 增加网络层数，总能减小训练集错误率
- B. 减小网络层数，总能减小测试集错误率
- C. 增加网络层数，可能增加测试集错误率
- C. 增加神经网络层数，确实可能提高模型的泛化性能，但不能绝对地说更深的网络能带来更小的错误率，还是要根据实际应用来判断，比如会导致过拟合等问题，因此只能选 C。

7. 说明 L_p 范数间的区别

L_1 范数：向量中各个元素绝对值之和

L_2 范数：向量中各个元素平方和的开二次方根

L_p 范数：向量中各个元素绝对值的 p 次方和的开 p 次方根

8. 用梯度下降训练神经网络的参数，为什么参数有时会被训练为 nan 值？

输入数据本身存在 nan 值，或者梯度爆炸了（可以降低学习率、或者设置梯度的阈值）

9. 卷积神经网络 CNN 中池化层有什么作用？

减小图像尺寸即数据降维，缓解过拟合，保持一定程度的旋转和平移不变性。

10. 请列举几种常见的激活函数。激活函数有什么作用？

sigmoi, relu, tanh。非线性化

11. 神经网络中 Dropout 的作用？具体是怎么实现的？

防止过拟合。。每次训练，都对每个神经网络单元，按一定概率临时丢弃。

12. 利用梯度下降法训练神经网络，发现模型 loss 不变，可能有哪些问题？怎么解决？

很有可能是梯度消失了，它表示神经网络迭代更新时，有些权值不更新的现象。

改变激活函数，改变权值的初始化等。

13. 如何解决不平衡数据集的分类问题？

可以扩充数据集，对数据重新采样，改变评价指标等。

14. 残差网络为什么能做到很深层？

15. 相比sigmoid激活函数ReLU激活函数有什么优势？

16. 卷积神经网络中膨胀卷积的作用是什么？

17. 解释下卷积神经网络中感受野的概念？如何计算感受野？如何增大感受野？
18. 模型欠拟合什么情况下会出现？有什么解决方案？
19. LSTM的效果为何比RNN好？
20. Caffe、TensorFlow、MxNet等框架的优缺点比较？
21. 多任务学习中标签缺失如何处理？
22. 梯度爆炸的解决方法？
23. 深度学习模型参数初始化都有哪些方法？
24. 注意力机制在深度学习中的作用是什么？有哪些场景会使用？
25. 卷积神经网络为什么会具有平移等不变性？
26. 神经网络中共享参数是指什么？哪些场景会使用？
27. 如何提高小型网络的精度？
28. 什么是神经网络的梯度消失问题，为什么会有梯度消失问题？有什么办法能缓解梯度消失问题？
在反向传播算法计算每一层的误差项的时候，需要乘以本层激活函数的导数值，如果导数值接近于 0，则多次乘积之后误差项会趋向于 0，而参数的梯度值通过误差项计算，这会导致参数的梯度值接近于 0，无法用梯度下降法来有效的更新参数的值。
改进激活函数，选用更不容易饱和的函数，如 ReLU 函数。
29. 列举你所知道的神经网络中使用的损失函数
欧氏距离，交叉熵，对比损失，合页损失
30. 对于多分类问题，为什么神经网络一般使用交叉熵而不用欧氏距离损失？

交叉熵在一般情况下更容易收敛到一个更好的解。

31. 1×1 卷积有什么用途？

通道降维，保证卷积神经网络可以接受任何尺寸的输入数据

32. 随机梯度下降法，在每次迭代时能保证目标函数值一定下降吗？为什么？

不能，每次迭代时目标函数不一样

33. 梯度下降法，为什么需要设置一个学习率？

使得迭代之后的值在上次值的邻域内，保证可以忽略泰勒展开中的二次及二次以上的项

34. 解释梯度下降法中动量项的作用

利用之前迭代时的梯度值，减小震荡

35. 为什么现在倾向于用小尺寸的卷积核？

用多个小卷积核串联可以有和大卷积核同样的能力，而且参数更少，另外有更多次的激活函数作用，增强非线性

36. 解释 GoogLeNet 的 Inception 模块的原理

对输入图像用多个不同尺寸的卷积核、池化操作进行同时处理，然后将输出结果按照通道拼接起来

37. 解释反卷积的原理和用途

反卷积即转置卷积，正向传播时乘以卷积核的转置矩阵，反向传播时乘以卷积核矩阵由卷积输出结果近似重构输入数据，上采样

38. 解释批量归一化的原理

在数据送入神经网络的某一层进行处理之前，对数据做归一化。按照训练样本的批量进行处理，先减掉这批样本的均值，然后除以标准差，然后进行缩放和平移。缩放和平移参数同训练得到。预测时使用训练时确定的这些值来计算

39. 解释 SVM 核函数的原理

核函数将数据映射到更高维的空间后处理，但不用做这种显式映射，而是先对两个样本向量做内积，然后用核函数映射。这等价于先进行映射，然后再做内积。

40. 什么是过拟合，过拟合产生的原因是什么？有什么方法能减轻过拟合？

过拟合指在训练集上表现的很好，但在测试集上表现很差，推广泛化能力差。产生过拟合的原因是训练样本的抽样误差，训练时拟合了这种误差。增加训练样本，尤其是样本的代表性；正则化

41. 什么样的函数可以用作激活函数？

非线性，几乎处处可到，单调

42. 什么是鞍点问题？

梯度为 0, Hessian 矩阵不定的点, 不是极值点

43. 在训练深度神经网络的过程中, 遇到过哪些问题, 怎么解决的?

不收敛, 收敛太慢, 泛化能力差。调整网络结构, 调整样本, 调整学习率, 调整参数初始化策略

44. SVM 如何解决多分类问题

多个二分类器组合。1 对 1 方案, 1 对剩余方案, 多类损失函数

45. 列举你知道的聚类算法

层次聚类, k 均值算法, DBSCAN 算法, OPTICS 算法, 谱聚类

46. K 均值算法中, 初始类中心怎么确定

随机选择 K 个样本作为类中心, 将样本随机划分成 K 个子集然后计算类中心

47. 简述 EM 算法的原理

EM 算法用于求解带有隐变量的最大似然估计问题。由于有隐变量的存在, 无法直接用最大似然估计求得对数似然函数极大值的公式解。此时通过 Jensen 不等式构造对数似然函数的下界函数, 然后优化下界函数, 再用估计出的参数值构造新的下界函数, 反复迭代直至收敛到局部极小值点



长按识别上方二维码关注