AIGC算法工程师面经—公式理解篇（上）

原创 喜欢瓦力的卷卷 瓦力算法学研所 2024年04月16日 16:11 广东

面试经验专栏

本篇总结了AIGC面经中最常问的公式理解类问题与相应答案。

本篇开始重点介绍面经中可能会问到的公式理解类题目及其答案。

在实际面试中，这类问题很大概率需要手写，或者需要很清晰地讲出公式含义及原理，这个过程中可能会遭到反复拷打，甚至手撕代码；因此本篇列出的问题只是一个方向会常问的问题，关于该方向更深入的问题会在问题下方列出。

在本篇中，为了便于辅助记忆与掌握原理，公众号文章中尽量用简洁的文字介绍，并提供了附有问题类型总结、详细公式与原理讲解的ppt版本便于大家后续复习。

下面是一个问题的快捷目录，需要注意答案中每个问题下方可能有引申的更深入的问题。

面试题

1. 手写softmax公式，手写BN公式，softmax层的label 是什么 ？

2. 交叉熵公式，分类为什么用交叉熵不用平方差 ？

3. 手推lr梯度, 交叉熵损失为什么有log项？为什么取负 ？

4. NER任务的损失函数是什么，写出来并解释一下?

5. 为什么逻辑回归用sigmoid激活函数？多分类逻辑回归是否也是sigmoid?

6. KL loss 公式是什么？一般什么情况下用的?

7.请解释ReLU激活函数的公式以及它的作用?

8.在卷积神经网络中，卷积操作的数学表达式是什么？请解释卷积核、步长和填充在其中的作用?

9.请解释残差连接的公式和原理，并说明它在深度神经网络中的作用?

10.请解释LSTM中遗忘门的计算公式以及其作用?

11.请解释L1和L2正则化的公式，以及它们在深度学习中的作用。

12.什么是贝叶斯定理？请写出其数学表达式，并解释每个部分的含义。

答案

1. **手写softmax公式，手写BN公式，softmax层的label 是什么 ？**

Softmax简介

Softmax 是一个常用于多类别分类任务中的激活函数，它将原始的分类得分转换为概率分布。Softmax 函数作为神经网络输出层的一部分，将神经网络的输出转换为各个类别的概率分布。

Softmax核心

最大概率的标签，能够将未规范化的预测变换为非负数，并且总和为1，同时能够让模型保持可导。

$$a_k = g(z_k) = \frac{e^{z_k}}{\sum_{i=1}^{C} e^{z_i}}$$

其中，Softmax的公式如上所示，g(·)代表激活函数，C代表类别，通常也指最后一层网络层的输出节点个数，Z指节点的输出值

1）手写softmax公式:

$$\text{softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}}$$

2）手写BN公式:

$$\text{BN}(x_i) = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \times \gamma + \beta$$

3）Softmax层的label是指训练样本的真实类别标签。

备注：其实还可能会有其他延申的问题，包括softmax缺点，使用时候的注意事项等，具体推荐大家去看看ppt。

问题2：Softmax缺点是什么？使用Softmax时需要注意什么问题?

1.对大量类别的计算开销： Softmax 函数涉及到对所有类别得分进行指数运算和求和，这在类别数量较大时会导致计算复杂度增加，影响模型的训练和推理速度。

2.梯度爆炸和消失：在反向传播过程中，由于 Softmax 函数的导数具有指数形式，当输入的值较大时，Softmax 的导数可能会变得非常小/大，从而导致梯度消失问题/爆炸。

因此需要注意：

1.缩放输入数值范围： Softmax 函数的输出受到输入值的依赖性较大，因为 Softmax 的计算依赖于所有输入的指数函数。如果不缩放，容易产生偏好数值较大的特征。

2. **交叉熵公式，分类为什么用交叉熵不用平方差 ?请用代码实现多分类交叉熵。**

交叉熵（Cross-Entropy）损失函数的公式通常用于衡量两个概率分布之间的差异，特别是在分类任务中常被使用。对于二分类问题，其公式如下：

$$H(y, \hat{y}) = -\left(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})\right)$$

对于多分类，公式如下：

$$H(y,\hat{y}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

分类问题使用交叉熵而不是平方差的原因是：

1）由于分类问题输出层常常要经过softmax，使用MSE后对损失函数求梯度，其中包含对softmax层的求导，不仅可能会导致错误的梯度（为0），也会在0、1处导致梯度很小（以二分类为例，此时softmax退化为sigmoid）

2）计算损失函数时，交叉熵损失只与真实类别对应输出有关，而MSE还需要计算其他错误类别的输出的平方。而当错误类别的输出都相等时，这部分值最小。因此MSE引入了一个关于类别信息的先验：各个类别之间的关系是等同的。

代码实现：

```
import numpy as np

def cross_entropy_loss(y, y_hat):
    """

    参数：
    y : 实际标签，
    y_hat : 模型预测概率

    返回值：
    loss : 交叉熵损失
    """
    epsilon = 1e-10
    # 计算交叉熵损失
    loss = -np.sum(y * np.log(y_hat + epsilon)) / len(y)
    return loss

y_true = np.array([[1, 0, 0], [0, 1, 0], [0, 0, 1]])
y_pred = np.array([[0.9, 0.05, 0.05], [0.05, 0.89, 0.06], [0.05, 0.01, 0.94]]
# 计算交叉熵损失
loss = cross_entropy_loss(y_true, y_pred)
print("交叉熵损失：", loss)

```

备注：这里也容易出现一些引申问题，也是具体推荐大家看一下ppt。

- 交叉熵和KL loss是什么关系
- 基于熟悉的框架（pytorch）写一段代码运用交叉熵求导的二分类代码

问题1：交叉熵与相对熵(KL散度)的关系

对于训练数据分布A（标签的分布）和模型输出分布B之间的KL散度可以用一下公式表示：

$$D_{KL}(A\|B)=\sum_i p_A(v_i)\log p_A(v_i)-p_A(v_i)\log p_B(v_i)$$

即A和B的KL散度=A的熵-AB的交叉熵，在机器学习中，训练数据的分布是固定的，因此最大化相对熵（KL散度）等价于最小化交叉熵，也等价于极大似然估计。

3. **手推lr梯度，交叉熵损失为什么有log项？为什么取负？**

对于交叉熵损失函数，LR的梯度计算为:

$$\frac{\partial J}{\partial w_i}=(y-\hat{y})x_i$$

交叉熵损失中的log项是为了惩罚模型对真实标签的错误预测，取负是因为优化问题中通常是最小化损失函数。

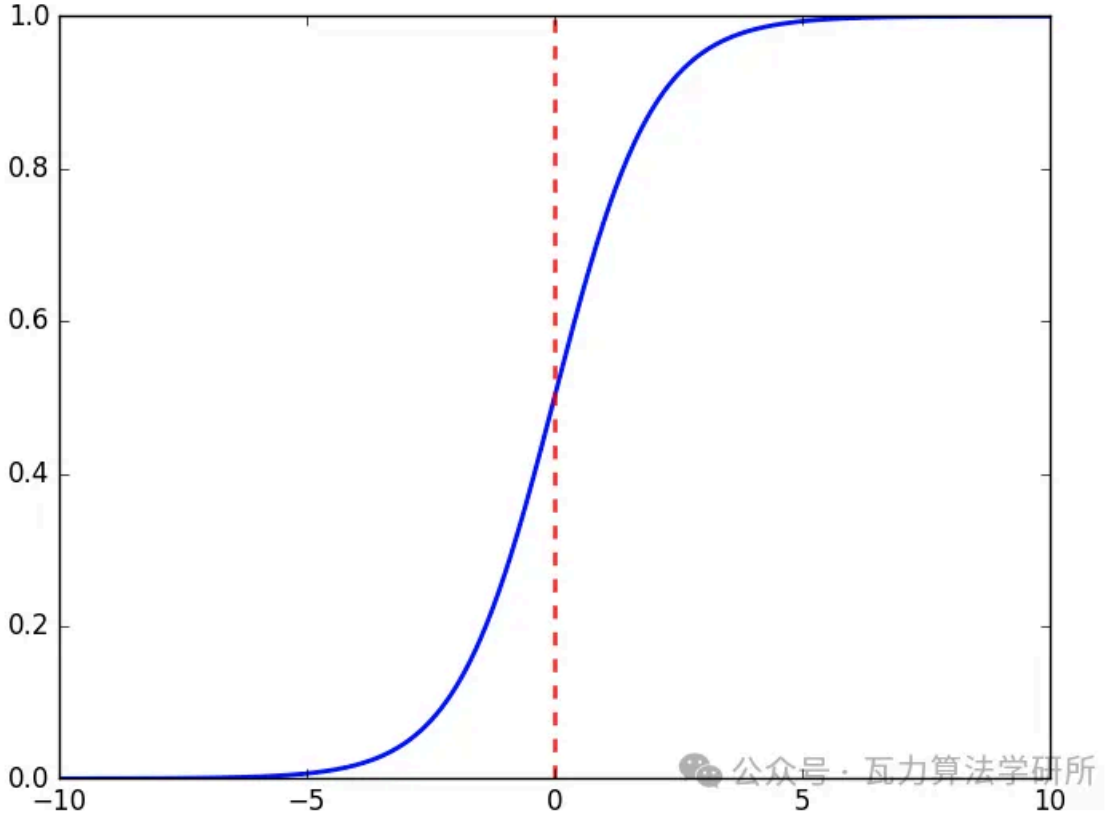
4. **NER任务的损失函数是什么，写出来并解释一下?**

NER损失函数通常是条件随机场（CRF）损失函数。它的核心思想公式如下：

$$L=-\log P(Y|X)$$

这个损失函数的含义是最大化给定输入序列 X 下真实标签序列 Y 的条件概率，即最大化预测的准确性。**由于是对每个token进行分类预测，会结合多分类交叉熵损失一起使用。**

5. **为什么逻辑回归用sigmoid激活函数？多分类逻辑回归是否也是sigmoid?**



sigmoid函数图像如上所示。

逻辑回归用sigmoid激活函数是因为其输出在 0 到 1 之间，可以被解释为样本属于某一类别的概率。在多分类逻辑回归中，通常会使用softmax激活函数，因为它可以将模型的输出转化为各个类别的概率分布。

6. **KL loss 公式是什么？一般什么情况下用的?**

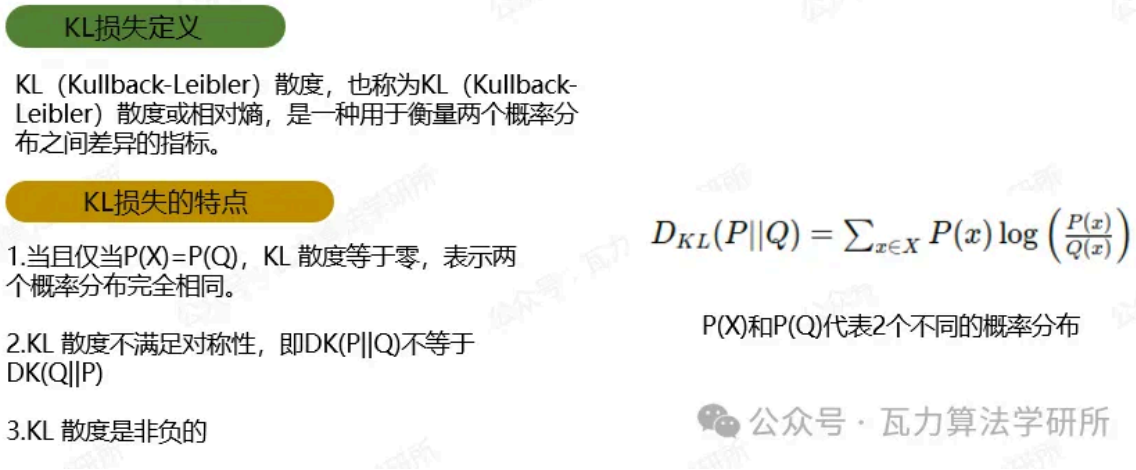
KL（Kullback-Leibler）散度用于衡量两个概率分布的差异。其公式为：

$$D_{KL}(P||Q) = \sum_i P(i) \log\left(\frac{P(i)}{Q(i)}\right)$$

其中，P 和 Q 是两个概率分布。KL散度通常在训练生成模型时用作正则化项，帮助模型生成更接近真实分布的样本。

备注：这里也容易出现一些引申问题，ppt中均有详细总结。

- 如何减小两个概率分布之间的KL散度？讲讲优化方法
- KL散度是否对称？如果不对称，如何解释其不对称性？
- KL散度有什么局限性？在什么情况下不适合使用？



剩下的6个问题将在后续公众号中持续推送~

想要获取资料的同学请关注公众号，并且回复“公式理解”获取完整版本ppt。

瓦力算法学研所
我们是一个致力于分享人工智能、机器学习和数据科学方面理论与应用知识的公众号。我...
117篇原创内容

公众号

面试干货 70

面试干货 · 目录

上一篇
AIGC算法工程师面经——模型训练通识基础篇

下一篇
大模型面经——LLama2和chatGLM相对于transformer具体做了哪些优化?