LLMs 损失函数篇

来自： AiGC面试宝典

宁静致远

2024年01月28日 13:20

一、介绍一下 KL 散度?

KL（Kullback-Leibler）散度衡量了两个概率分布之间的差异。其公式为：

$$D_{KL}(P//Q) = -\sum_{x \in X} P(x) \log \frac{1}{P(x)} + \sum_{x \in X} P(x) \log \frac{1}{Q(x)}$$

二、交叉熵损失函数写一下，物理意义是什么?

交叉熵损失函数（Cross-Entropy Loss Function）是用于度量两个概率分布之间的差异的一种损失函数。在分类问题中，它通常用于衡量模型的预测分布与实际标签分布之间的差异。

$$H(p,q) = -\sum_{i=1}^{N} p_i \log(q_i) - (1-p_i) \log(1-q_i)$$

> 注：其中，p 表示真实标签，q 表示模型预测的标签，N 表示样本数量。该公式可以看作是一个基于概率分布的比较方式，即将真实标签看做一个概率分布，将模型预测的标签也看做一个概率分布，然后计算它们之间的交叉熵。
> 物理意义：交叉熵损失函数可以用来衡量实际标签分布与模型预测分布之间的“信息差”。当两个分布完全一致时，交叉熵损失为0，表示模型的预测与实际情况完全吻合。当两个分布之间存在差异时，损失函数的值会增加，表示预测错误程度的大小。

三、KL 散度与交叉熵的区别?

KL散度指的是相对熵，KL散度是两个概率分布P和Q差别的非对称性的度量。KL散度越小表示两个分布越接近。也就是说KL散度是不对称的，且KL散度的值是非负数。（也就是熵和交叉熵的差）

- 交叉熵损失函数是二分类问题中最常用的损失函数，由于其定义出于信息学的角度，可以泛化到多分类问题中。
- KL散度是一种用于衡量两个分布之间差异的指标，交叉熵损失函数是KL散度的一种特殊形式。在二分类问题中，交叉熵函数只有一项，而在多分类问题中有多项。

四、多任务学习各loss差异过大怎样处理?

多任务学习中，如果各任务的损失差异过大，可以通过动态调整损失权重、使用任务特定的损失函数、改变模型架构或引入正则化等方法来处理。目标是平衡各任务的贡献，以便更好地训练模型。

五、分类问题为什么用交叉熵损失函数不用均方误差（MSE）?

交叉熵损失函数通常在分类问题中使用，而均方误差（MSE）损失函数通常用于回归问题。这是因为分类问题和回归问题具有不同的特点和需求。
分类问题的目标是将输入样本分到不同的类别中，输出为类别的概率分布。交叉熵损失函数可以度量两个概率分布之间的差异，使得模型更好地拟合真实的类别分布。它对概率的细微差异更敏感，可以更好地区分不同的类别。此外，交叉熵损失函数在梯度计算时具有较好的数学性质，有助于更稳定地进行模型优化。
相比之下，均方误差（MSE）损失函数更适用于回归问题，其中目标是预测连续数值而不是类别。MSE损失函数度量预测值与真实值之间的差异的平方，适用于连续数值的回归问题。在分类问题中使用MSE损失函数可能不太合适，因为它对概率的微小差异不够敏感，而且在分类问题中通常需要使用激活函数（如sigmoid或softmax）将输出映射到概率空间，使得MSE的数学性质不再适用。
综上所述，交叉熵损失函数更适合分类问题，而MSE损失函数更适合回归问题。

六、什么是信息增益?

信息增益是在决策树算法中用于选择最佳特征的一种评价指标。在决策树的生成过程中，选择最佳特征来进行节点的分裂是关键步骤之一，信息增益可以帮助确定最佳特征。
信息增益衡量了在特征已知的情况下，将样本集合划分成不同类别的纯度提升程度。它基于信息论的概念，使用熵来度量样本集合的不确定性。具体而言，信息增益是原始集合的熵与特定特征下的条件熵之间的差异。
在决策树的生成过程中，选择具有最大信息增益的特征作为当前节点的分裂标准，可以将样本划分为更加纯净的子节点。信息增益越大，意味着使用该特征进行划分可以更好地减少样本集合的不确定性，提高分类的准确性。

七、多分类的分类损失函数(Softmax)?

多分类的分类损失函数采用Softmax交叉熵（Softmax Cross Entropy）损失函数。Softmax函数可以将输出值归一化为概率分布，用于多分类问题的输出层。Softmax交叉熵损失函数可以写成：

$$-\sum_{i=1}^{n} y_i \log(p_i)$$

> 注：其中，n是类别数，yi是第i类的真实标签，pi是第i类的预测概率。

八、softmax和交叉熵损失怎么计算，二值交叉熵呢?

softmax计算公式如下：

$$y = \frac{e^{f_i}}{\sum_j e^{fj}}$$

多分类交叉熵：

$$L = \frac{1}{N}\sum_i L_i = -\frac{1}{N}\sum_i \sum_{c=1}^{M} y_{ic} \log$$

其中：

(p_{ic})

- M ——类别的数量
- y_{ic} 一符号函数（0 或 1），如果样本 i 的真实类别等于 c 取 1，否则取 0
- p_{ic} 一观测样本 i 属于类别 c 的预测概率

二分类交叉熵：

$$L = \frac{1}{N}\sum_i L_i = \frac{1}{N}\sum_i$$

$$-[y_i \cdot \log(p_i) + (1-y_i) \cdot \log(1-p_i)]$$

- y_i — 表示样本 i 的label, 正类为 1，负类为 0
- p_i — 表示样本 i 预测为正类的概率

九、如果softmax的e次方超过float的值了怎么办？

将分子分母同时除以 x 中的最大值，可以解决。

$$\tilde{x}_k = \frac{e^{xk-\max(x)}}{e^{x_1-\max(x)}+e^{x_2-\max(x)}+\ldots+e^{x_k-\max(x)}+\ldots+e^{x_n-\max(x)}}$$