

Cross Entropy Loss

简介

作用

交叉熵是用来衡量两个概率分布相似性

联系

经常将softmax与 cross-entropy放在一起讨论,目前来看主要是有以下几点原因:

1. softmax 可以将原始输出值(logits/vector)转换为概率分布,可以作为cross-entropy的输入
2. 数学形式和梯度计算上具有高度适配性

公式

Softmax

$$p_j = \frac{e^{z_j}}{\sum_{j=1}^K e^{z_j}}$$

- 输入: $z = [z_1, z_2, \dots, z_j, \dots, z_K]$, 其中K是类别数
- 输出: 归一化概率值 $p = [p_1, p_2, \dots, p_j, \dots, p_K]$, 其中 $\sum_{j=1}^K p_j = 1$

CrossEntropyLoss

$$Loss = - \sum_{j=1}^K y_j * \log(p_j)$$

- $y = [y_1, y_2, \dots, y_j, \dots, y_K]$ 表示真实分布, 通常是one-hot编码(如 $y = [0, 1, 0]$)
- p表示概率值, 可以认为是上述 softmax的输出

二分类交叉熵损失

$$BinaryCEL = - \frac{1}{N} \sum_{i=1}^N [y_i * \log(p_i) + (1 - y_i) * \log(1 - p_i)]$$

- y_i 表示真实标签
- p_i 表示模型预测的概率值
- N 表示样本数

多分类交叉熵损失

$$CEL = - \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K y_{i,k} * \log(p_{i,k})$$

- $y_{i,k}$ 表示第i个样本属于第k类的真是标签
- $p_{i,k}$ 表示模型预测的第i个样本属于第k类的概率值
- N 表示样本数

- K表示类别数

参考

[softmax和交叉熵的数学联系](#)

[小白都能看懂的softmax详解](#)

[The Softmax function and its derivative](#)

[一文看懂softmax loss](#)