

成本(损失)函数

2023年4月4日 上午 07:44

成本函数 (Cost Function)，也称**损失函数 (Loss function)**用来度量模型的预测值 $f(x)$ 与真实值 Y 的差异程度的运算函数，它是一个非负实值函数，通常使用 $L(Y, f(x))$ 来表示。

1. 在回归问题中，常用的成本函数是均方误差 (Mean Squared Error · MSE)，其计算公式为：

$$J(w, b) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$$

2. 在分类问题中，常用的成本函数是交叉熵 (Cross Entropy)，其计算公式为：

$$J(w, b) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log(h_w(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_w(x^{(i)}))]$$

损失函数使用主要是在模型的训练阶段，每个批次的训练数据送入模型后，通过前向传播输出预测值，然后损失函数会计算出预测值和真实值之间的差异值，也就是损失值。得到损失值之后，模型通过反向传播去更新各个参数，来降低真实值与预测值之间的损失，使得模型生成的预测值往真实值方向靠拢，从而达到学习的目的。

作用：将输出与基本事实进行比较

损失函数的梯度作用：知道应该移动多少来更新，以此让损失下降，让该函数更好地识别猫。反复做此操作，直到找到合适的参数插入架构

如何选择损失函数？

通常情况下，损失函数的选取应从以下方面考虑：

- (1) 选择最能表达数据的主要特征来构建基于距离或基于概率分布度量的特征空间。
- (2) 选择合理的特征归一化方法，使特征向量转换后仍能保持原来数据的核心内容。
- (3) 选取合理的损失函数，在实验的基础上，依据损失不断调整模型的参数，使其尽可能实现类别区分。
- (4) 合理组合不同的损失函数，发挥每个损失函数的优点，使它们能更好地度量样本间的相似性。
- (5) 将数据的主要特征嵌入损失函数，提升基于特定任务的模型预测精确度。

目标检测用YOLO损失函数：

式中对宽高都进行开根是为了使用大小差别比较大的边界框差别减小。例如，一个同样将一个100x100的目标与一个10x10的目标都预测大了10个像素，预测框为110 x 110与20 x 20。显然第一种情况我们还可以失道接受，但第二种情况相当于把边界框预测大了一倍，但如果不使用根号函数，那么损失相同，都为200。但把宽高都增加根号时：

$$(\text{sqrt}(20) - \text{sqrt}(10))^2 = 3.43$$

$$(\text{sqrt}(110) - \text{sqrt}(100))^2 = 0.48$$

显然，对小框预测偏差10个像素带来了更高的损失。通过增加根号，使得预测相同偏差与更小的框产生更大的损失。