损失函数

## 对数损失函数

对数损失函数（LogLoss）是经常在离线评估中使用的指标，在一个二分类问题中，LogLoss的定义如下：

其中，为输入实例的真实类别，为预测输入实例是正样本的概率，N是样本总数。

## 交叉熵损失函数

熵(Entropy)源于信息学，用于度量信息的不确定度。熵越大，代表信息的不确定性越大，信息也就越大（一种类比的理解就是，对于十分稳定的物体来说，一直不发生变化是确定的，发生变化所蕴含的信息量就更大）。某个分布P(i)的熵定义为：

实际上，𝐻(𝑃)也可以使用其他底数的log函数计算。举个例子，对于 4 分类问题，如果某个样本的真实标签是第4类，那么标签的One-hot编码为[0,0,0,1]，即这张图片的分类是唯一确定的，它属于第4类的概率𝑃(𝑦为4|𝒙) = 1，不确定性为0，它的熵可以简单的计算为：

也就是说，对于确定的分布，熵为0，不确定性最低。如果它预测的概率分布是[0.1,0.1,0.1,0.7]，它的熵可以计算为：

这种情况比前面确定性类别的例子的确定性要稍微大点。

考虑随机分类器，它每个类别的预测概率是均等的：[0.25,0.25,0.25,0.25]，同样的方法，可以计算它的熵约为2，这种情况的不确定性略大于上面一种情况。

由于𝑃(𝑖) ∈ [0,1], log2 𝑃(𝑖) ≤ 0，因此熵𝐻(𝑃)总是大于等于0。当熵取得最小值0时， 不确定性为0。

交叉熵(Cross Entropy)的定义如下：

𝐻(𝑝||𝑞) ≜ −∑𝑝(𝑖)log2 𝑞(𝑖)

参考：

龙书，TensorFlow深度学习。