

# 损失函数的原理和类别

什么是损失函数？

- 损失函数是关于模型输出与label的函数，用来衡量模型在训练集上的表现效果，它是模型对数据拟合程度的反映、拟合得越差，损失函数的值就应该越大。

Loss Function?Cost Function?Object Function? 分不清楚？

- Loss Function：是单个样例的损失/误差
- Cost Function：是对数据集整体的误差描述，是选定参数w和b后对数据进行估计所要支付的代价，cost是对多有数据的误差取平均值得到的。
- Object Function（目标函数）定义为：Cost Function+正则化项

## 常见的损失函数

- 0-1损失函数
- 平均损失函数
- 绝对值损失函数
- 对数损失函数
- 全局损失函数
- 交叉熵损失函数

### 0-1损失函数（0-1 loss function）

$$L(Y, f(x)) = \begin{cases} 1, Y \neq f(x) \\ 0, Y = f(x) \end{cases}$$

### 平均损失函数（quadratic loss function）

$$L(Y, f(x)) = (Y - f(x))^2$$

### 绝对值损失函数（quadratic loss function）

$$L(Y, f(x)) = |Y - f(x)|$$

### 对数损失函数（logarithmic loss function）

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

### 全局损失函数

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

对于平方损失函数，为了方便求导，我们可以在前面乘上一个 $\frac{1}{2}$ ，和平方项求导后的2抵消，即：

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

### 交叉熵损失函数（cross entropy loss function）

- 熵： $H(y) = -\sum_i y_i \log(y_i)$
- 交叉熵： $H(y_i a) = -\sum_i y_i \log(a_i)$
- 平均交叉熵： $H(y_i a) = -\frac{1}{n} \sum_n \sum_i y_{i,n} \log(a_{i,n})$

二分类cross entropy（逻辑回归的损失函数）：

$$H(y, a) = -\frac{1}{n} \sum_n y \log(a) + (1 - y) \log(1 - a)$$

多分类的cross entropy

$$-\sum_{c=1}^M \log(P_{o,c})$$

- M-类别的数量；
- y-指示变量（0或1），如果该类别和样本观测到的类别相同就是1，否则是0；
- p-对于观测样本属于类别c的预测概率

## 计算方法

- 神经网络最后一层得到每一个类别的得分scores
- 该得分经过softmax转换为概率输出
- 模型预测的类别概率输出与真实类别的one hot形式进行cross entropy损失数的计算

为什么交叉熵可以衡量模型的准确率？

- 信息论的角度来看，交叉熵等价于信赖出来的模型（分布）与真实模型（分布）之间的分布差异（两者相差一个只和样本拿数据量有关的倍数N），而这个交叉熵的大小，衡量了训练模型与真实模型之间的差距，交叉熵越小，从而说明模型越准确。

相对于平方损失函数的优势？

- 平方损失函数： $c = \frac{1}{2}(a - y)^2$
- 参数更新： $\frac{\delta C}{\delta W} = (a - y)\sigma'(a)x^T$   
 $\frac{\delta C}{\delta b} = (a - y)\sigma'(a)$
- 交叉熵的参数更新  
 $H' = \frac{1}{n} \sum (a_n - y_n) = \frac{1}{n} \sum (\sigma(z_n) - y_n)$
- 当误差打的时候，权重更新就快；当误差小的时候，权重更新就慢

如何选择损失函数？

Task	Output layer activation	Loss function	Expression of loss function
Regression	Linear	Mean squeraed error(MSE)	$L(y, t) =   y - t  _2^2$
Binary classification	Sigmoid	Binary cross-entropy	$L(y, t) = -\sum_{i=1}^n \log y_i$ $= -\sum_{i=1}^n (1 - t_i) \log(1 - y_i)$
Multi-class classification	Softmax	Categorical cross-entropy	$L(y, t) = -\sum_{i=1}^n t_i \log y_i$

结合激活函数来选择

- Sigmoid 系意外的激活函数+MSE
- Sigmoid + Cross Entropy
- Softmax + Cross Entropy/log-likelihood

其他损失函数

- Adaboost的指数损失函数
- Central loss等
- 根据实际场景可以自行设定不同的损失函数