损失函数

损失函数（loss function）是用来估量你模型的预测值f(x)与真实值Y的不一致程度，它是一个非负实值函数,通常使用L(Y, f(x))来表示，损失函数越小，模型的鲁棒性（稳定性）就越好。损失函数是经验风险函数的核心部分，也是结构风险函数重要组成部分。模型的结构风险函数包括了经验风险项和正则项，通常可以表示成如下式子：



其中，前面的均值函数表示的是经验风险函数，L代表的是损失函数，后面的是正则化项，或者叫惩罚项，它可以是L1，也可以是L2，或者其他的正则函数。整个式子表示的意思是找到使目标函数最小时的值。

经验风险

模型f(x)关于训练数据集的平均损失称为经验风险或经验损失记作



在假设空间、损失函数以及训练数据集确定的情况下，经验风险函数式就可以确定。经验风险最小化(ERM)的策略认为，经验风险最小的模型是最优的模型，所以经验最小化模型就是求解最优化问题。



其中F是假设空间

当样本容量足够大时，经验风险最小化可以保证有很好的学习效果，但是样本容量很小时，效果未必很好，会产生过拟合现象。结构风险最小化(SRM)是为了防止过拟合而提出的。结构风险最小化等价于正则化。结构风险在经验风险上加上表示模型复杂度的正则化项或罚项，在假设空间、损失函数以及训练数据集确定的情况下，结构风险最小化定义为



其中 为模型的复杂度,是定义在假设空间F上的泛函.模型f越复杂,复杂度就越大; 反之, 模型f越简单, 复杂度就越小. 也就是说, 复杂度表示了对复杂模型的惩罚.λ>=0是系数,用以权衡经验风险和模型复杂度.结构风险小需要经验风险与模型复杂度同时小.结构风险小的模型往往对训练数据以及未知的测试数据都有较好的预测.

结构风险最小化(SRM)的策略认为，结构风险最小的模型是最优的模型，所以结构最小化模型就是求解最优化问题。

+

这样，监督学习问题就变成了经验风险或结构风险函数的最优化问题。

下面介绍几种损失函数：

(1)0-1损失函数



(2)绝对值损失函数



(3)平方损失函数（最小二乘法，Ordinary Least Squares）



最小二乘法是线性回归的一种，OLS将问题转化成了一个凸优化问题。在线性回归中，它假设样本和噪声都服从高斯分布（为什么假设成高斯分布呢？其实这里隐藏了一个小知识点，就是中心极限定理），最后通过极大似然估计（MLE）可以推导出最小二乘式子。最小二乘的基本原则是：最优拟合直线应该是使各点到回归直线的距离和最小的直线，即平方和最小。换言之，OLS是基于距离的，而这个距离就是我们用的最多的欧几里得距离。为什么它会选择使用欧式距离作为误差度量呢（即Mean squared error， MSE），主要有以下几个原因：

a.简单，计算方便；

b.欧氏距离是一种很好的相似性度量标准；

c.在不同的表示域变换后特征性质不变。

Y-f(X)表示的是残差，整个式子表示的是残差的平方和，而我们的目的就是最小化这个目标函数值（注：该式子未加入正则项），也就是最小化残差的平方和（residual sum of squares，RSS）。

而在实际应用中，通常会使用均方差（标准差）作为一项衡量指标，公式如下：

上面提到了线性回归，这里额外补充一句，我们通常说的线性有两种情况，一种是因变量y是自变量x的线性函数，一种是因变量y是参数的线性函数。在机器学习中，通常指的都是后一种情况。

（最小二乘法推导可以参考以下链接<https://blog.csdn.net/lql0716/article/details/70165695>）

(4)对数损失函数（逻辑回归）



对数损失, 即对数似然损失(Log-likelihood Loss), 也称逻辑斯谛回归损失(Logistic Loss)或交叉熵损失(cross-entropy Loss), 是在概率估计上定义的.它常用于(multi-nominal, 多项)逻辑斯谛回归和神经网络,以及一些期望极大算法的变体. 可用于评估分类器的概率输出.

对数损失通过惩罚错误的分类,实现对分类器的准确度(Accuracy)的量化. 最小化对数损失基本等价于最大化分类器的准确度.为了计算对数损失, 分类器必须提供对输入的所属的每个类别的概率值, 不只是最可能的类别. 对数损失函数的计算公式如下:

https://images2018.cnblogs.com/blog/757205/201806/757205-20180623163230097-750693662.png

其中, Y 为输出变量, X为输入变量, L 为损失函数. N为输入样本量, M为可能的类别数, 是一个二值指标, 表示类别 j 是否是输入实例 的真实类别. 为模型或分类器预测输入实例 属于类别 j 的概率.

如果只有两类 {0, 1}, 则对数损失函数的公式简化为

https://images2018.cnblogs.com/blog/757205/201806/757205-20180623170758959-855776533.png

这时, yi 为输入实例 xi 的真实类别, pi 为预测输入实例 xi 属于类别 1 的概率. 对所有样本的对数损失表示对每个样本的对数损失的平均值, 对于完美的分类器, 对数损失为 0

在逻辑回归的推导中，它假设样本服从伯努利分布（0-1分布），然后求得满足该分布的似然函数，接着取对数求极值等等。而逻辑回归并没有求似然函数的极值，而是把极大化当做一种思想，进而推导出它的经验风险函数为：最小化负的似然函数（即max F(y, f(x)) —-> min -F(y, f(x)))。从损失函数的视角来看，它就成了log损失函数了。

（5）Hinge损失函数

在机器学习算法中，hinge损失函数和SVM是息息相关的。在线性支持向量机中，最优化问题可以等价于下列式子

$$\min_{w,b}  \ \sum_{i}^{N} [1 - y_i(w\cdot x_i + b)]_{+} + \lambda||w||^2 $$

下面对式子变形，令

$$[1 - y_i(w\cdot x_i + b)]_{+} = \xi_{i}$$

于是原式就变成了

$$\min_{w,b}  \ \sum_{i}^{N} \xi_i + \lambda||w||^2 $$

若取C，式子可变成

$$\min_{w,b}  \frac{1}{C}\left ( \frac{1}{2}\ ||w||^2 $$ + C \sum_{i}^{N} \xi_i\right )$$

可看出，该式子与下式费非常相似：

$$\frac{1}{m} \sum_{i=1}^{m} l(w \cdot  x_i + b, y_i) + ||w||^2$$

前半部分中的就是hinge损失函数，而后面相当于L2正则项。

在libsvm中一共有4中核函数可以选择，对应的是-t参数分别是：

0-线性核；

1-多项式核；

2-RBF核；

3-sigmoid核。