# 机器学习之损失函数详解

## 机器学习和大数据的渊源

大数据4V特征：

(1)数据量大：TP--PB--ZB（大数据中的框架可解决这个问题）

(2)数据的多样化：（大数据框架解决了该问题）

结构化数据----类似于Excel或Mysql中存储的

半结构化数据--Html或Xml

非结构化数据--视频、音频等

(3)速度快：（大数据框架解决了该问题）

数据的增长速度快

数据的处理速度快--Hadoop--Storm--Spark--Flink

(4)价值密度低或价值高：（机器学习解决）

价值密度的衡量=有价值的数据/整体数据

价值密度问题是大数据平台无法解决的，需要结合机器学习去解决

## 2.为什么要有损失函数

机器学习主要是研究如何使计算机从给定的数据中学习规律，即从观测数据（样本）中寻找规律，并利用学习到的规律（模型）对未知或无法预测的数据进行预测。

损失函数是衡量决策函数好坏而定义的。

给定一个实例(x,y)，真实目标是y，机器学习模型的预测为f(x,theta)。如果预测错误时（f(x,theta)!=y），我们常常需要定义一个度量函数来定量地计算错误的程度。常见的损失函数有01损失函数、平方损失、指数损失等。

## 3.分类问题损失函数

### 3.1 0-1损失函数

**【分类问题（目标函数1）】**



### 3.2 0-1损失的Sklearn实现

Zero\_one\_loss 函数计算n个样本上的0-1分类损失的和或者平均值，默认情况下是所有样本的损失的平均损失，把参数normalize设置为False就可以返回损失之和。

官网提供的参考代码如下



### 3.3交叉熵损失函数

（负对数似然损失函数）：【分类问题】

**让去估计给定x的情况下第i个类别的条件概率。**【**分类问题**2】。

**对于分类问题，预测目标y为离散的类别，模型输出f(x,theta)为每个类的条件概率。**

假设，模型预测的第i个类的条件概率，则f(x,theta)满足：



可以看作真实类别y的似然函数。参数可以直接用最大似然估计来优化。考虑到计算问题，我们经常使用最小化负对数似然，也就是负对数似然损失函数：

**【分类问题】**

如果我们用one-hot向量y来表示目标类别c，其中只有，其余的向量元素都为0。

负对数似然函数也可以写为：

**【分类问题】**

yi也可以看成是真实类别的分布，上面公式正是交叉熵的形式。因此，**负对数似然损失函数**也常叫做**交叉熵损失函数**，上式也是负对数似然函数的一种改进。

注：在数字电路中，one-hot是一种状态编码，指对任意给定的状态，状态寄存器只有1位为1，其余都是0.

### 3.4交叉熵损失实现

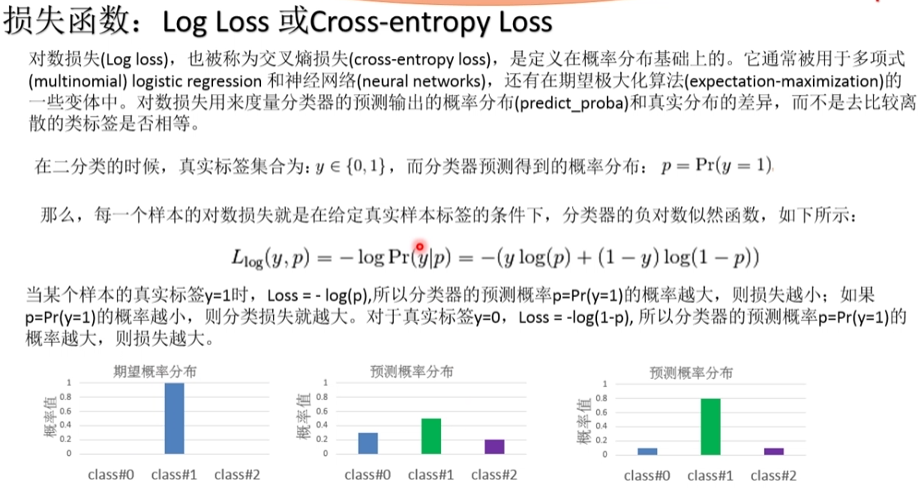
交叉熵损失和其他的损失函数区别就在于交叉熵损失是定义在概率分布基础上的，它通常是被用于基于多项式的逻辑斯特回归和神经网络中，还有用在EM算法中。对数损失是用来度量分类器的预测输出的概率分布和真是分布的差距，而不是比较离散类标签是否相等。

在二分类的时候，真是标签集合为{0,1},而分类器预测得到的概率分布：P(y=1)

则定义样本的对数损失就是在给定真是样本标签下，分类器的负对数似然函数

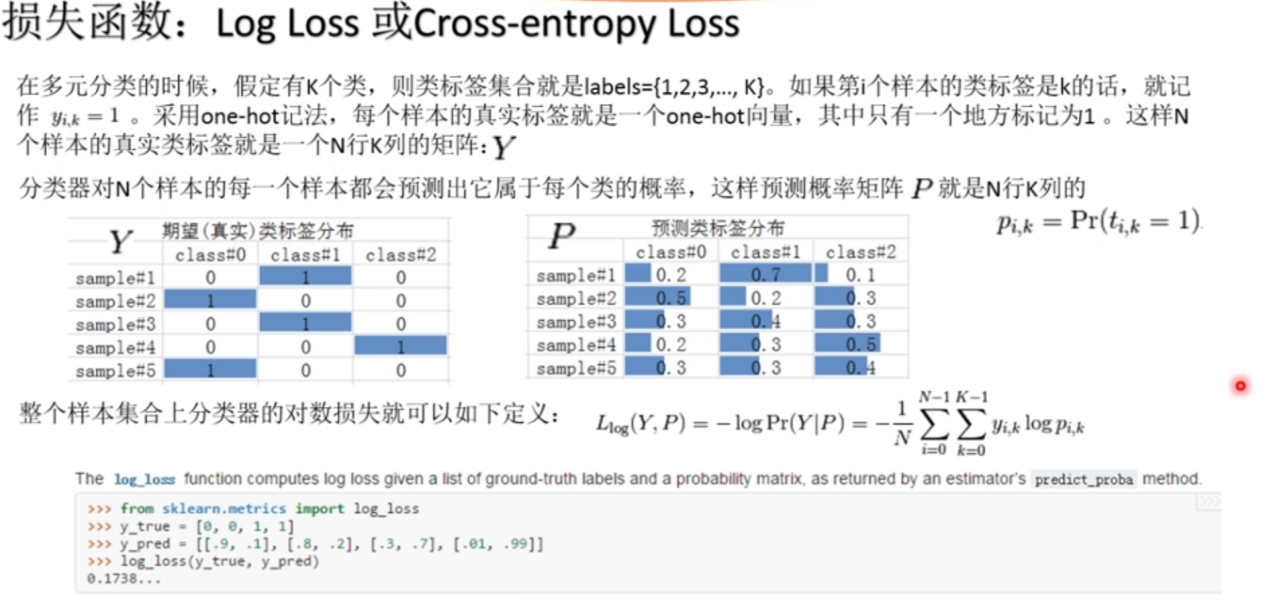


当某个样本的真实标签为y=1的时候，loss=-log(p)，所以分类器的预测概率P(y=1)的概率越大，加个符号就得到损失越小；如果P(y=1)概率越小，则分类损失就越大。



### 3.5交叉熵损失多分类实现

多元分类的时候，假定有K个类，则类标签集合就是labels={1,2,3,...k}。如果第i个样本的类标签就是k的话，就记为y=1，这时采用上述提到的one-hot编码，就可以使得只有一个地方标记为1.这样N个样本的真实标签就是一个N行K列的矩阵Y：

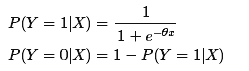


### 3.6交叉熵损失推导

Logistic Regression 采用的则是负 log 损失，

L(Y,P(Y|X))=−logP(Y|X)

从二项分布的角度来考虑 Logistic 回归：



这里另 z=θTx ,  δ 为 sigmod 映射，则：

   E(z)=−log(δ(z))

E(z)的图形如下图的红色曲线，可见 z 越接近 1 ， E(z) 的取值越小，即损失越小。反之另：

E(z)=1−log(δ(z))

此时得到的图像应该为关于 E(z)对称的红色的线（没画出），此时 z越接近 -1，E(z) 的取值越小，即损失越小。

### 3.7 Hinge损失函数及Sklearn实现

Hinge\_loss函数计算模型与数据之间的平均距离，是一个仅仅考虑预测误差单边度量，通常用于最大间隔分类器的SVM。

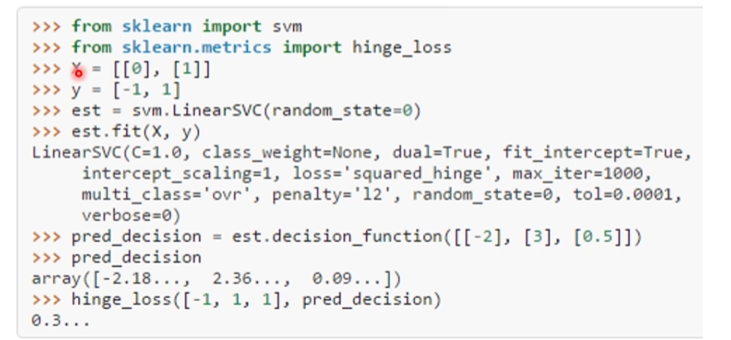
下式中y是真正的标签，w是decision\_function输出的预测到的边界，hinge损失定义：



如果是多个标签，hinge\_loss函数语句Crammer&Singer提出的多分类决策办法来计算。如果yw是对真实类标签的预测，并且yt是对多有其他类标签的预测的最大值，多类别标签的损失如下：



代码实现：



官网文档展示了hine\_loss函数度量SVM分类器在多元分类问题中的使用方法：



### 3.8Hinge损失的二分类问题

SVM 求解使通过建立二次规划原始问题，引入拉格朗日乘子法，然后转换成对偶的形式去求解，这是一种理论非常充实的解法。这里换一种角度来思考，在机器学习领域，一般的做法是经验风险最小化 ERM ，即构建假设函数为输入输出间的映射，然后采用损失函数来衡量模型的优劣。求得使损失最小化的模型即为最优的假设函数，采用不同的损失函数也会得到不同的机器学习算法，比如这里的主题 SVM 采用的是 Hinge Loss 。

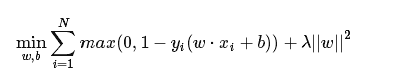
给定数据集  ， 要用这些数据做一个线性分类器，即求得最优分离超平面 wx+b=0来将样本分为正负两类，给定数据集后只需求得最优的参数  w,b即可，为了解决这个问题，首先做出如下线性映射函数

y=wx+b

根据经验风险最小化原则， 这里引入二分类的 Hinge Loss :

max(0,1−yi(wxi+b))

   上图中对应的 E(z)=max(0,1−z) ，所以SVM可以通过直接最小化如下损失函数二求得最优的分离超平面：



### 3.9Hinge损失的多分类问题

对于多分类问题，现在要用这些数据做一个 k 类的线性分类器 ,现在需要优化的参数变为 W,b， 此时的 W∈Rk×n，为一个 k×n的矩阵，b∈Rk为一个向量，现在的映射关系如下 ：s=Wxi+b，此时有 s∈Rk ，s 中的每个分量代表分类器在该类别的得分，样本 xi的标签  yi∈Rk , 这里若 xi属于类别 k，则 yi中除了第 k个分量外其余元素全为 0 ，比如 5 分类问题， xi 属于第 3 类，则有  yi=[0,0,1,0,0] , 用 sj表示得分向量 s 中的第 j个分量 ， Syi 表示对应 yi=1的分量，则单个样本多分类的Hinge Loss可表示为：



所以 k分类线性分类SVM 的 Hinge Loss表示为：



## 4.其他损失函数

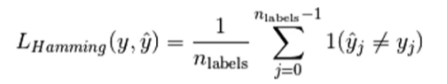
### 4.1平方损失函数

**【回归问题】**



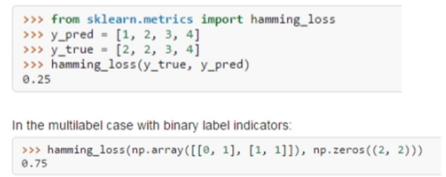
### 4.2汉明损失及代码实战

Hamming\_loss计算两个样本集合之间的平均汉明距离



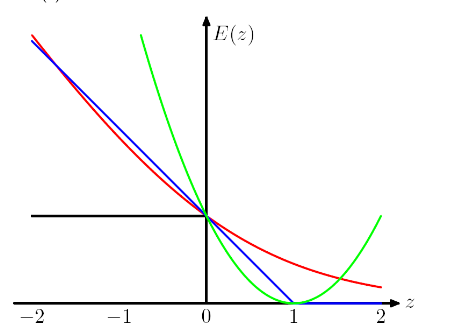
假定yi1是对给定样本的第j个标签的预测值，yi是此样本的真实标签，是类的数量或者标签的数量。

官网提供参考代码如下，最下面是多分类的情况。



## 总结

通过图示在认识损失函数：



**注**： 图中绿色的线为 square loss ，蓝色的线为 hinge loss， 红的的线为负 log 损失。

损失函数汇总：

