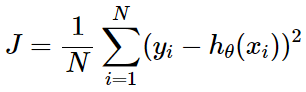
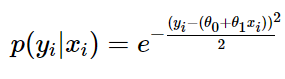
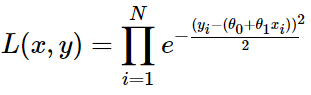
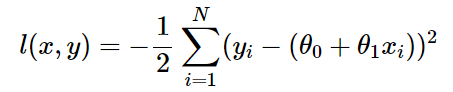
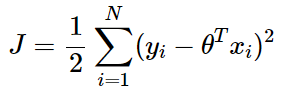
**Loss Function**

https://rohanvarma.me/Loss-Functions/

**Mean Square Error均方误差**：常用于回归问题。形式如。以下为用经典线性回归问题对MSE进行推导：

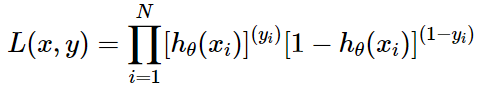
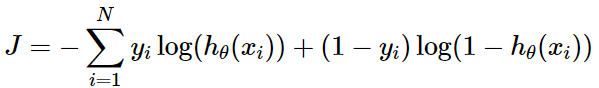
，其中η为噪声，服从μ=0，σ=1的高斯分布。则Y的期望和方差为：，，得出一个观察样本的概率分布为：

，假设所有样本各自独立，在观察N个样本后的似然函数为：

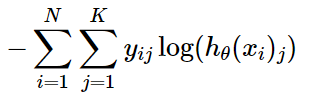
，取对数后为：，从而可得到MSE为：。从推导过程可以看出，如果假设输出（带有不可忽略的高斯噪声，常量mean和variance）是输入的实数函数，采用MSE是一个合理的选择。如果假设不成立，如分类问题（不是高斯分布，因为分布函数不是连续的），MSE就可能不是最好的选择。

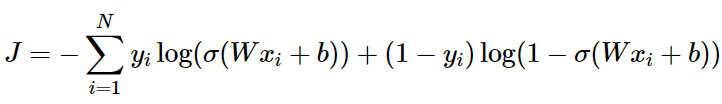
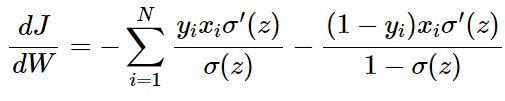
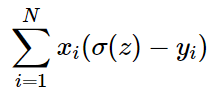
**Cross-Entropy交叉熵（二元分类）**：模型预测结果h(θ)被激活函数σ(wx+b)-softmax转化为0-1区间的概率，大于0.5的为正样本，小于0.5的为负样本。因此，概率表达为：，。

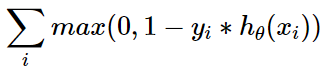
假设数据为IID（independent and identically distributed），则似然函数为：

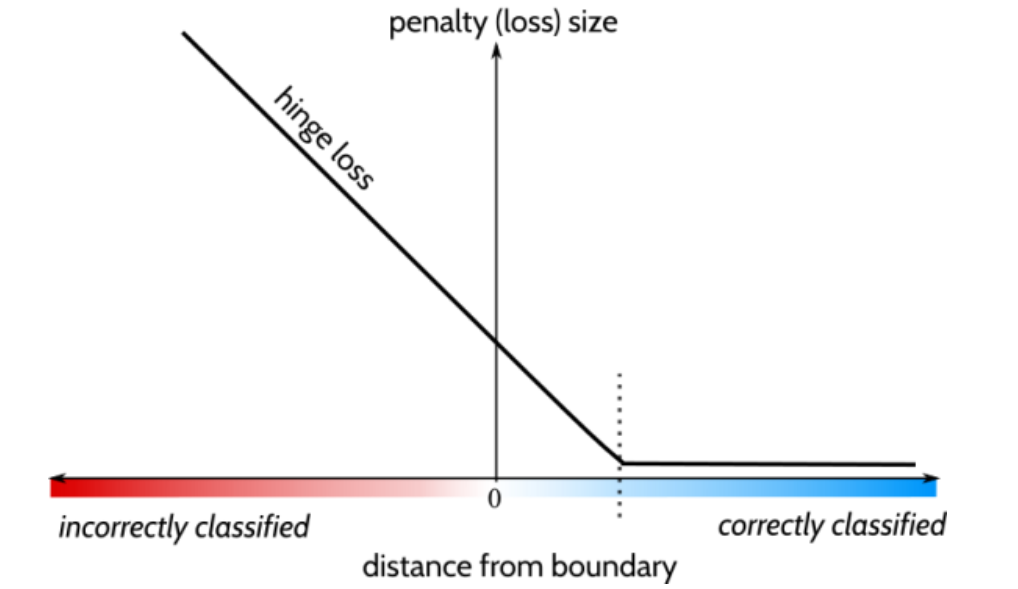
，取对数：

**Cross-Entropy交叉熵（多元分类）**：对于多类别情况，分类是不连续的，如果用数字代表每个类别，那将会很高效，但是在分类器中往往会把数据认为是连续有序的，但实际上多元分类中不是有序的，而是随机分配的，所以不能直接用数字代表各个类别。因此在多元分类中，采用one-hot encoding来表示分类结果Y，其方法是使用N位状态寄存器来对N个状态进行编码，每个状态都有它独立的寄存器位，并且在任意时候，其中只有一位有效。通俗点理解就是每个状态有一堆0，只有一个为1，1的位置就是激活的位置，表示该特定的类别。交叉熵公式则为：

，其中N表示样本数，K表示类别数。则特定样本的损失为

**Cross-Entropy交叉熵（信息论）**：在信息论中，测量两个概率分布差别有多大，即 KL发散程度。可认为分类问题有两个概率分布：（1）真实样本的概率分布，所有概率质量都集中在正确标签上，没有概率质量分布在其他标签；（2）模型预测的概率分布，概率密度集中度是由softmax函数得到的输出。假设在分类任务中采用MSE和sigmoid激活函数，在梯度下降时，求得的梯度带有sigmoid导数项，而sigmoid导数分布很接近0。在随机初始化权重时，初始化值是接近0的，这表明在训练初期因为权重太小就会很快使得梯度变小，在模型还没训练好的情况下就产生梯度消失，同时训练初期的梯度下降效率也会很低，因为权重减少的幅度不大。如果采用的是交叉熵和sigmoid，则，，简化后可得，可以看出这个结果里面不包含σ的导数项，而主要取决于预测结果和真实值的误差，即σ-y。

**(L2-regularized) hinge loss：**常用于SVM，像一个铰链合页形状，与ReLu方向相反，为0的一侧为正确样本，因此惩罚值为0。斜率一侧主要惩罚惩罚不正确或正确但是程度不太高的样本。对于二元分类：。其中y可以是-1，1。只有当|h|>=1时才会无惩罚。举例：（1）当h为0.2，y为-1时，惩罚值则为1.2，因为0.2距离-1太远了，所以惩罚值很大；（2）当h为-0.7，y为-1时，惩罚值为0.3，因为-0.7距离-1很近，所以惩罚之小；（3）当h为-1.1时，y为-1，惩罚值为-0.1，去较大值0，所以没有惩罚。



**Hinge loss和cross-entropy的区别**：前者试图最大化margin以确保每个点是正确可分的。后者是最大化模型参数组成的似然函数。在交叉熵中，采用了从softmax得到的概率分数，以此根据分数来理解各个模型参数中的相互概率分布。

对于回归问题，常见的损失函数有MSE(L2 loss, L2 loss也常常作为正则项。当预测值与目标值相差很大时, 梯度容易爆炸，因为梯度里包含了x−t)，RMSE，MAE(L1 loss，由于L1 loss具有稀疏性，为了惩罚较大的值，因此常常将其作为正则项添加到其他loss中作为约束。L1 loss的最大问题是梯度在零点不平滑，导致会跳过极小值)，R-square。 对于分类问题，常见的损失函数有softmax loss(softmax with cross-entropy loss)及其变种、KL散度，hinge loss，指数loss，逻辑loss