1. 0-1损失函数

> 感知机 / 朴素贝叶斯

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

2. 绝对值损失函数

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

3. 均方误差损失 MSE

> 线性回归

$$L(Y,F(X))=(Y-f(X))^2$$

4. 对数损失函数

> 逻辑斯蒂回归

$$L(Y,f(x))=-log(P(y|x))$$

如果是二分类$y\in\{0,1\}$，则 $P(y|x)=p(y_i=1|x_i)^{y_i}p(y_i=0|x_i)^{1-y^i}$

所以：

$$L(Y,f(x))=-log(p(y_i=1|x_i)^{y_i}p(y_i=0|x_i)^{1-y^i})=-(y_i logp(y_i=1|x)+(1-y_i)logp(y_i=0|x))$$

对于二分类来说，$p(y_i=0|x_i)=1-p(y_i=1|x_i)$

所以, 设$p(y_i=1|x_i)=p$:

$$L(Y,f(x))=-(y_i logp+(1-y_i)log(1-p))$$

等价于二分类交叉熵损失

5. 交叉熵损失 CE

二分类交叉熵损失

> 等价于对数损失函数

$$L(Y,f(x))=-(y_i logp+(1-y_i)log(1-p))$$

常用在sigmoid后，进行损失函数计算（每个label是独立分布的）

多分类交叉熵损失

$L(Y,f(x))=-\sum y_i logp_i$

常用在softmax后，进行损失函数计算

6. 合页损失 Hinge

> SVM

普通版：假设$Y\in\{1,-1\}$，如下公式表示，如何分类正确，损失为0，如果分类错误，损失为 $1-Yf(X)$

$L(Y,f(X))=max(0,1-Yf(X))$

优化版: 目的是让正得分与负得分的差距大于1

$L(Y,f(X))=max(0,1+f(y=0|x)-f(y=1|x))$

7. 均方误差（MSE）和交叉熵（CE）损失对比

- **MSE无差别得关注全部类别上预测概率和真实概率的差.交叉熵关注的是正确类别的预测概率.**

  举个例子， 实际标签是$[1,0,0]$,模型预测得到的概率是$[0.9,0.4,0.3]$,那么交叉熵损失函数的结果是$1log(0.9)+0log(0.4)+0log(0.3)$,可以看到，后面两项的系数都为0，可以直接忽略。而MSE则都得全部算一遍
- **交叉熵更有利于梯度更新**

  MSE会收敛的慢一些，因为它求导的结果相比于交叉熵还多乘以一个sigmod函数。如果当前模型的输出接近0或者1时，梯度很小。
- **MSE是假设数据符合高斯分布时,模型概率分布的负条件对数似然;交叉熵是假设模型分布为多项式分布时,模型分布的负条件对数似然。**