损失函数.

1. 分类错误率 .

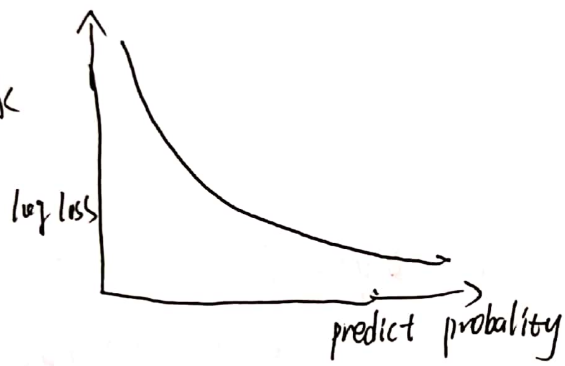
2. 均方误差 $MSE = \frac{1}{n}\sum_{i}^{n}(\hat{y}_i - y_i)^2$

3. 交叉熵 : $L = \frac{1}{N}\sum_i L_i = \frac{1}{N}\sum_i -[y_i \log(p_i) + (1-y_i)\cdot\log(1-p_i)]$

y_i 为 label 正类为1 负类为0

二、多分类 $L = \frac{1}{N}\sum_i L_i = -\frac{1}{N}\sum_i \sum_{c=1}^{M} y_{ic}\log(p_{ic})$

图像



学习过程.

w. scores

$\rightarrow s_i = 1.8 \rightarrow \sigma(s_i) = \frac{1}{1+e^{-s_i}} \rightarrow p_i = 0.14 \rightarrow L_i = -[y_i \cdot \log p_i + (1-y_i)\log(1-$

$\frac{\partial s_i}{\partial w_i}$ $\frac{\partial p_i}{\partial s_i}$ $\frac{\partial L_i}{\partial p_i}$

$$\frac{\partial L_i}{\partial w_i} = \frac{1}{N}\frac{\partial L_i}{\partial w_i} = \frac{1}{N}\frac{\partial L_i}{\partial p_i}\cdot\frac{\partial p_i}{\partial s_i}\frac{\partial s_i}{\partial w_i}$$

$\frac{\partial L_i}{\partial p_i}$ $L_i = -[y_i \log(p_i) + (1-y_i)\log(1-p_i)]$

$$= \frac{\partial -[y_i \log(p_i) + (1-y_i)\log(1-p_i)]}{\partial p_i} = -\frac{y_i}{p_i} - [(1-y_i)\cdot\frac{1}{1-p_i}\cdot(-1)]$$

$$= -\frac{y_i}{p_i} + \frac{1-y_i}{1-p_i}$$

$$p = \sigma(s) = \frac{1}{1+e^{-x}} \quad \frac{\partial p_i}{\partial s_i} = \sigma(s_i)(1-\sigma(s_i))$$

$$\frac{\partial s_i}{\partial w_i} = x_i$$

$$\frac{\partial l_i}{\partial w_i} = \frac{\partial l_i}{\partial p_i} \cdot \frac{\partial p_i}{\partial s_i} \cdot \frac{\partial s_i}{\partial w_i} = \left[-\frac{y_i}{p_i} + \frac{1-y_i}{1-p_i}\right] \cdot \sigma(s_i) \cdot [1-\sigma(s_i)] \cdot x_i$$

$$= [\underbrace{\sigma(s_i) - y_i}] \cdot x_i$$

$$\cancel{FL(p_t) = -\alpha_t (1-p_t)^{\gamma} \log(p_t)}$$

Focal loss.

$$FL(p_t) = -(1-p_t)^{\overset{1}{\gamma}} \overset{0.7}{\log}(p_t) \qquad CE(p,y) = \begin{cases} -\log(p) & \text{if } y=1 \\ -\log(1-p) & \text{otherwise.} \end{cases}$$

CE 简单负样本过多

1、当样本 $y=1$ 是 true label 样本被分错 $(1-p_t)$ 接近1 损失不受影响

$p_t \to 1$ $(1-p_t)$ 接近0 分的好 样本权值调低、降低分得好对 loss 样本的贡献