# 类别不平衡损失函数

# 样本不均衡问题

- 假设样本不均衡，正样本数远小于负样本数；
- 分类损失函数：

二分类交叉熵：$$L = \frac{1}{N}\sum_i L_i = \frac{1}{N}\sum_i -[y_i \cdot log(p_i) + (1-y_i)\cdot log(1-p_i)]$$

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

N：表示为样本数

负样本数量太大，将占总的loss的大部分

# 交叉熵&加权交叉熵

交叉熵损失：

$$\mathrm{CE}(p, y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1-p) & \text{otherwise.} \end{cases} \quad (1)$$

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{otherwise,} \end{cases} \quad (2)$$

$$\mathrm{CE}(p, y) = \mathrm{CE}(p_t) = -\log(p_t).$$

加权交叉熵损失：

$$\mathrm{CE}(p_t) = -\alpha_t \log(p_t). \quad (3)$$

通过设定a的值来控制正负样本对总的loss的共享权重。a取比较小的值来降低负样本（多的那类样本）的权重。

# Focal Loss

公式3虽然可以控制正负样本的权重，但是没法控制容易分类和难分类样本的权重；

Focal Loss 损失函数：

$$\mathrm{FL}(p_t) = -(1 - p_t)^{\gamma} \log(p_t). \qquad (4)$$

γ称作调制系数（modulating factor）

目的是通过减少易分类样本的权重，从而使得模型在训练时更专注于难分类的样本。

# Focal Loss

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise,} \end{cases} \quad (2)$$

- 若一个样本容易分错，则Pt较小，趋近于0，则（1-Pt）接近1，损失函数不受影响；
- 若一个样本较好分类（well-classified），则Pt趋近于1，（1-Pt）接近0，其权值降低；

- 当γ=0的时候，focal loss就是传统的交叉熵损失，当γ增加的时候，调制系数也会增加，专注参数γ平滑地调节了易分样本调低权值的比例。调制因子减少了易分样本的损失贡献，拓宽了样例接收到低损失的范围。

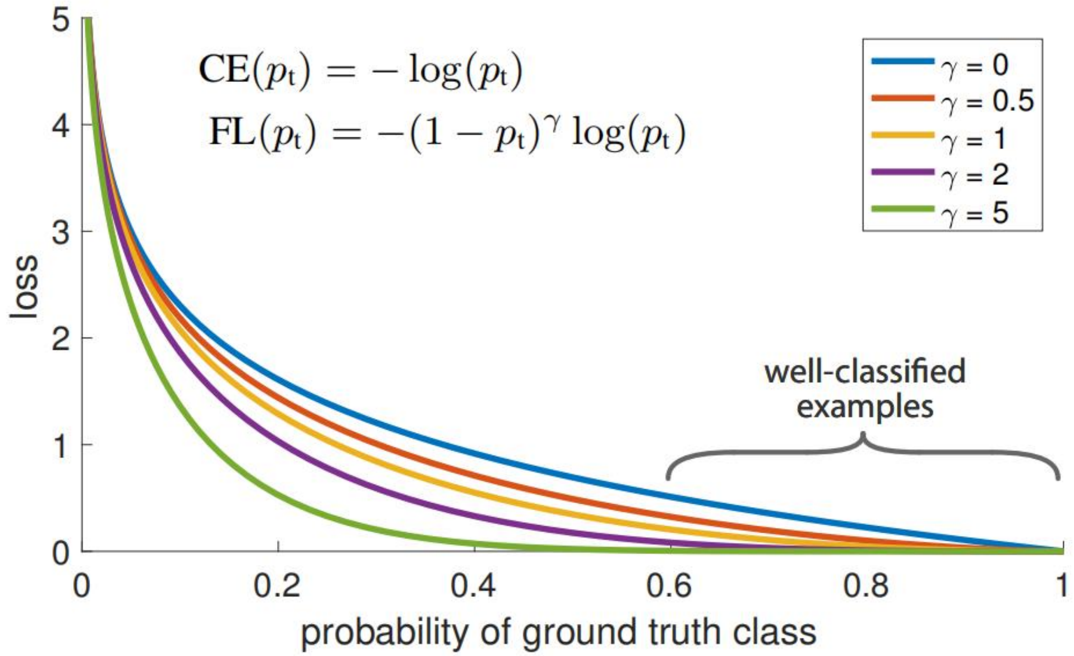



Figure 1. We propose a novel loss we term the *Focal Loss* that adds a factor $(1 - p_t)^{\gamma}$ to the standard cross entropy criterion. Setting $\gamma > 0$ reduces the relative loss for well-classified examples ($p_t > .5$), putting more focus on hard, misclassified examples. As our experiments will demonstrate, the proposed focal loss enables training highly accurate dense object detectors in the presence of vast numbers of easy background examples.

# Focal Loss

- 结合了公式3和公式4，这样既能调整正负样本的权重，又能控制难易分类样本的权重；
- 在实验中a的选择范围也很广，一般而言当γ增加的时候，a需要减小一点（实验中γ=2，a=0.25的效果最好）

$$\mathrm{CE}(p_t) = -\alpha_t \log(p_t). \tag{3}$$

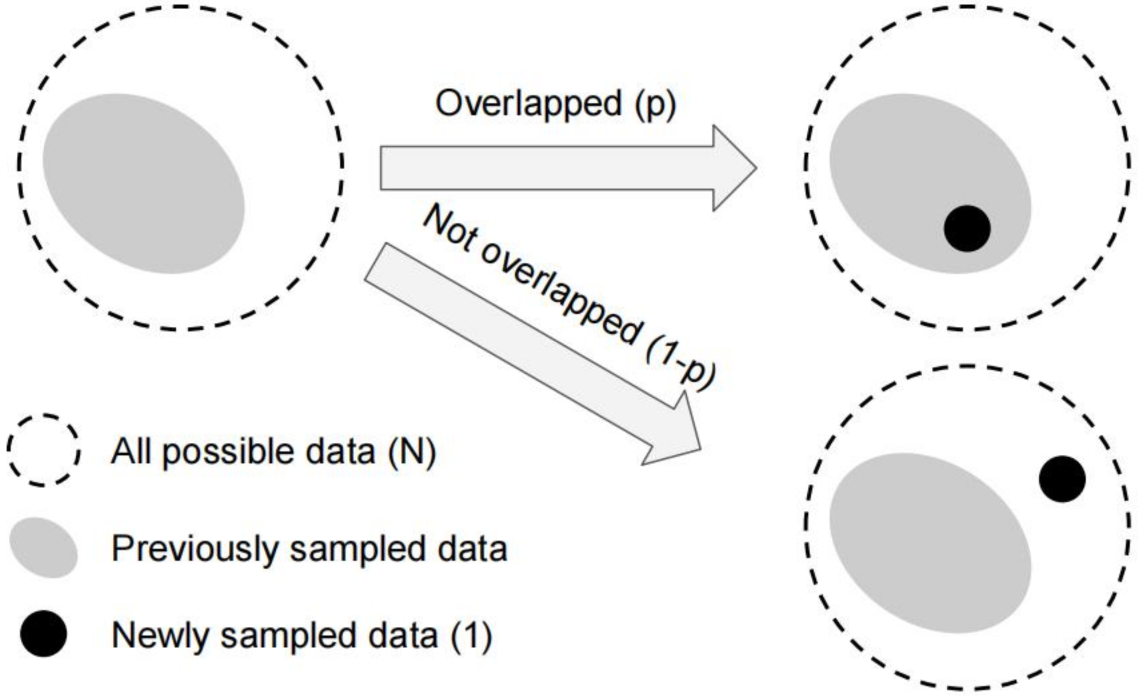
$$\mathrm{FL}(p_t) = -(1-p_t)^\gamma \log(p_t). \tag{4}$$

$$\mathrm{FL}(p_t) = -\alpha_t(1-p_t)^\gamma \log(p_t). \tag{5}$$

# CB Loss

- 重采样和基于样本数量的加权，在样本量达到一定时，增加新样本带来的收益会消失。

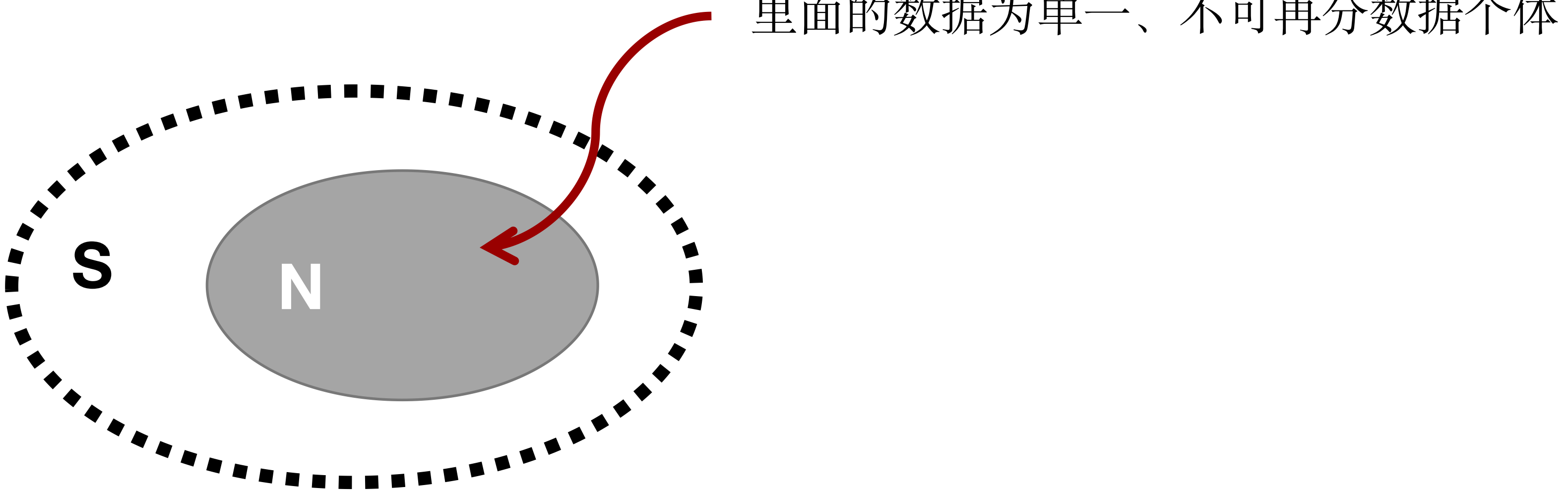
- 方法主要是通过引入一个加权因子，这个因子则和有效样本数成反比（通过有效样本数来平衡损失）




link: https://openaccess.thecvf.com/content_CVPR_2019/papers/Cui_Class-Balanced_Loss_Based_on_Effective_Number_of_Samples_CVPR_2019_paper.pdf

# CB Loss-计算有效样本数

1. 随机覆盖的数据采样方法




给定一个类别，定义这个类特征空间的所有可能的数据集合为$\mathcal{S}$。假设$\mathcal{S}$的体积为$N$并且$N \geqslant 1$。定义每一个数据都是$\mathcal{S}$的子集并且体积都是单位体积 `1` ，每个数据都有可能和其他数据重叠。考虑随机覆盖数据采样的过程，每个数据（也就是子集）都有可能被采样，最终目的是能够覆盖集合$\mathcal{S}$的全部可能。采样的数据越多，对集合$\mathcal{S}$的覆盖越好。采样数据的期望体积会随机数据量的增加而增加，最终的边界是$N$。

# CB Loss-计算有效样本数

2. 定义1：样本的有效数据量是样本的期望体积（expected volume）

为了保证这个问题是可解的，我们假设一个新的采样数据只能通过两种方式和之前采样的数据进行交互：

- 新样本存在于之前采样数据中的可能性为p（即：新样本是覆盖样本的概率）
- 新样本在之前采样数据之外的可能性为(1-p)（即：新样本不是覆盖样本的概率）

# CB Loss-计算有效样本数

3.

定义有效样本数为$E_n$，其中$n$表示样本的总数。显然，当样本总数只有1个时，$E_1 = 1$；现假设之前采样了$n-1$个样例，接着采样第$n$个样例，已采样数据的体积设为$E_{n-1}$，那么新采样数据有$p = \frac{E_{n-1}}{N}$概率和之前的样例重叠，所以经过第$n$次采样之后，期望体积为：

$$E_n = pE_{n-1} + (1-p)(E_{n-1}+1) = 1 + \frac{N-1}{N}E_{n-1}$$

现在假设$E_{n-1} = \frac{1-\beta^{n-1}}{1-\beta}$，那么：

$$E_n = 1 + \beta\frac{1-\beta^{n-1}}{1-\beta} = \frac{1-\beta^n}{1-\beta}$$

- 上式表明样例的有效数据量是总数$n$的指数函数，超参数$\beta \in [0,1)$控制$E_n$随着$n$的增长速度。

# CB Loss-类别平衡损失函数

（1）类别平衡损失用来处理不平衡数据集的问题，方法主要是**通过引入一个加权因子**，这个因子则**和有效样本数成反比。**

（2）假设输入样例$\mathbf{x}$的标签为$y \in \{1, 2, \cdots, C\}$，这里$C$表示类别总数。假设模型得到的概率分布$\mathbf{p} = [p_1, p_2, \cdots, p_C]^T$，这里$p_i \in [0, 1]$。定义损失函数为$\mathcal{L}(\mathbf{p}, y)$，假设类别$i$的样例数为$n_i$，根据上一节的公式可知对于类别$i$有$E_{n_i} = \frac{(1-\beta_i^{n_i})}{1-\beta_i}$，这里$\beta_i = \frac{N_i - 1}{N_i}$。

（3）为了平衡损失，引入权重因子$\alpha_i$，它和每个类的有效样本数成反比；为了保证总损失和之前大致一致，需要对权重因子归一化，得到$\sum_{i=1}^{C} \alpha_i = C$。为了简化，接下来使用$\frac{1}{E_{n_i}}$表示归一化的权重因子。

（4）对于一个类别为$i$的样例（对应的数量为$n_i$），在损失函数前加上权重因子$\frac{1-\beta}{1-\beta^{n_i}}$，其中$\beta \in [0, 1)$，那么类别平衡损失可以表示为：

$$\mathrm{CB}(\mathbf{p}, y) = \frac{1}{E_{n_y}} \mathcal{L}(\mathbf{p}, y) = \frac{1 - \beta}{1 - \beta^{n_y}} \mathcal{L}(\mathbf{p}, y)$$

- $n_y$表示训练集中类别为$y$的数量。

# CB Loss-应用三种损失函数

- 可用常用三种损失函数：$\mathcal{L}(\mathbf{p}, y)$

1. Softmax Cross-Entropy：

$$\mathrm{CB}_{\mathrm{softmax}}(\mathbf{z}, y) = -\frac{1-\beta}{1-\beta^{n_y}} \log\left(\frac{\exp(z_y)}{\sum_{j=1}^{C} \exp(z_j)}\right)$$

2. Sigmoid Cross-Entropy

$$z_i^t = \begin{cases} z_i, & \text{if } i = y \\ -z_i, & \text{otherwise} \end{cases}$$

$$\mathrm{CB}_{\mathrm{sigmoid}} = -\frac{1-\beta}{1-\beta^{n_y}} \sum_{i=1}^{C} \log\left(\frac{1}{1+\exp(-z_i^t)}\right)$$

3. Class-Balanced Focal Loss

$$\mathrm{FL}(\mathbf{z}, y) = -\sum_{i=1}^{C} (1-p_i^t)^{\gamma} \log(p_i^t)$$

$$\mathrm{CB}_{\mathrm{fl}}(\mathbf{z}, y) = -\frac{1-\beta}{1-\beta^{n_y}} \sum_{i=1}^{C} (1-p_i^t)^{\gamma} \log(p_i^t)$$

# 总结

- 样本不平衡如何影响损失函数
- 常用的三种不平衡条件下的损失函数
  - 加权交叉熵
  - Focal Loss
  - CB Loss