

# Face Recognition method survey:2

paper: ArcFace: Additive Angular Margin Loss for Deep Face [\[pdf\]](#)

**前言：**损失函数作为神经网络不可缺少的组成部分，在量化标签值与预测值之间的差距的同时，也引导着网络学习的特征空间的变化。一般来讲索引类任务，如人脸识别、行人重识别、图像检索学习到的子空间都要扩大类间间距同时缩小类内间距，本文是伦敦帝国理工学院邓健康等人在2018年发表的方法，在SphereFace的基础上改进了对特征想归一化和加性角度间隔，提高了类间可分性的同时加强了类内紧度和类间差异。

## 人脸识别的两种方法

使用深度卷积神经网络(DCNN)特征嵌入是人脸识别的首选方法，DCNN将人脸图像映射成特征向量表示，要求映射后的特征具有较小的类内距离，较大的类间距离。人脸识别的方法主要有两条研究主线：一种是他当成分类问题，在训练集上采用softmax损失函数训练，另一种是直接度量空间学习，比如triplet loss。

对于 softmax 损失的缺点：（1）学习到的特征对于闭集分类问题是可分离的，但对于开集人脸识别问题没有足够的辨别力；（2）计算量的大小随着身份数 N 线性增加。

对于三元组损失的缺点：（1）人脸三元组的数量存在组合爆炸，特别是对于大规模数据集，导致迭代步数显著增加；（2）semi-hard sample挖掘是一个非常困难的问题。

## 损失函数的演变

### Softmax

$$L_1 = -\frac{1}{m} \sum_{i=1}^m \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}}$$

上述公式中m代表了batch-size的大小，n代表了类别个数。Softmax损失函数并没有明确地扩大决策边界，即增加类间距离，减小类内距离。这也就为后面的改进留足了空间。

### Weights Normalization

如果将softmax中的偏置置0，对于向量内积部分  $W_j^T x_i$  部分，如果将权重归一化，即其L2范数为1，那么就会有原公式变换为  $L_2 = -\frac{1}{m} \sum_{i=1}^m \log \frac{e^{\|x_i\| \cos(\theta_{y_i})}}{e^{\|x_i\| \cos(\theta_{y_i})} + \sum_{j=1, j \neq y_i}^n e^{\|x_i\| \cos(\theta_{y_j})}}$ ，通过权重归一化得到的L2就只与特征向量和权重矩阵的夹角有关。

### Multiplicative Angular Margin

从上述的  $L_2$  入手，如果将角度通过乘以角度间隔m进行扩大，那么就得到了

$$L_3 = -\frac{1}{m} \sum_{i=1}^m \log \frac{e^{\|x_i\| \cos(m\theta_{y_i})}}{e^{\|x_i\| \cos(m\theta_{y_i})} + \sum_{j=1, j \neq y_i}^n e^{\|x_i\| \cos(m\theta_{y_j})}}。 \cos函数在(0,\pi)内是单调递减的，乘上正整数$$

m后内积会减小，这样可以加大类间的差别。由于Cosine函数不单调，通过一个分段函数  $\Psi(\theta_{y_i}) = (-1)^k \cos(m\theta_{y_i}) - 2k$  将损失函数转成单调的。

$$L_4 = -\frac{1}{m} \sum_{i=1}^m \log \frac{e^{\|x_i\| \psi(\theta_{y_i})}}{e^{\|x_i\| \psi(\theta_{y_i})} + \sum_{j=1, j \neq y_i}^n e^{\|x_i\| \cos(\theta_{y_j})}}$$

其中:  $\psi(\theta_{y_i}) = (-1)^k \cos(m\theta_{y_i}) - 2k, \theta_{y_i} \in [\frac{k\pi}{m}, \frac{(k+1)\pi}{m}], k \in [0, m-1], m \geq 1$

## Feature Normalization

特征和权重正则化消除了径向变化并且让每个特征都分布在超球面上, 将  $x$  的范数缩放为  $s$  的时候, sphereface loss 将会变为:  $L_5 = -\frac{1}{m} \sum_{i=1}^m \log \frac{e^{s\psi(\theta_{y_i})}}{e^{s\psi(\theta_{y_i})} + \sum_{j=1, j \neq y_i}^n e^{s\cos(\theta_{y_j})}}$

## CosFace

cosface 针对 sphereface 做了三点调整:

- loss 的形式做了稍微的改变, 将超参数  $m$  由乘法运算变成了减法运算
- 不仅对权重进行了正则化, 还对特征进行了正则化
- 再乘上一个  $s$  参数, 超球面过小时, 分类映射到超球面上不好分类, 这个  $s$  参数可以扩大超球面体积, 帮助分类

所以最终的损失函数变为了:

$$L_6 = \frac{1}{N} \sum_i -\log \left( \frac{e^{s(\cos(\theta_{y_i,i})-m)}}{e^{s(\cos(\theta_{y_i,i})-m)} + \sum_{j \neq y_i} e^{s\cos(\theta_{j,i})}} \right)$$

## Arcface

Arcface 从 CosFace 出发, 主要区别是将  $m$  放入了  $\cos$  中, 角度距离比余弦距离对角度的影响更加直接:

$$L_7 = \frac{1}{N} \sum_i -\log \left( \frac{e^{s(\cos(\theta_{y_i,i}+m))}}{e^{s(\cos(\theta_{y_i,i}+m))} + \sum_{j \neq y_i} e^{s\cos(\theta_{j,i})}} \right)$$

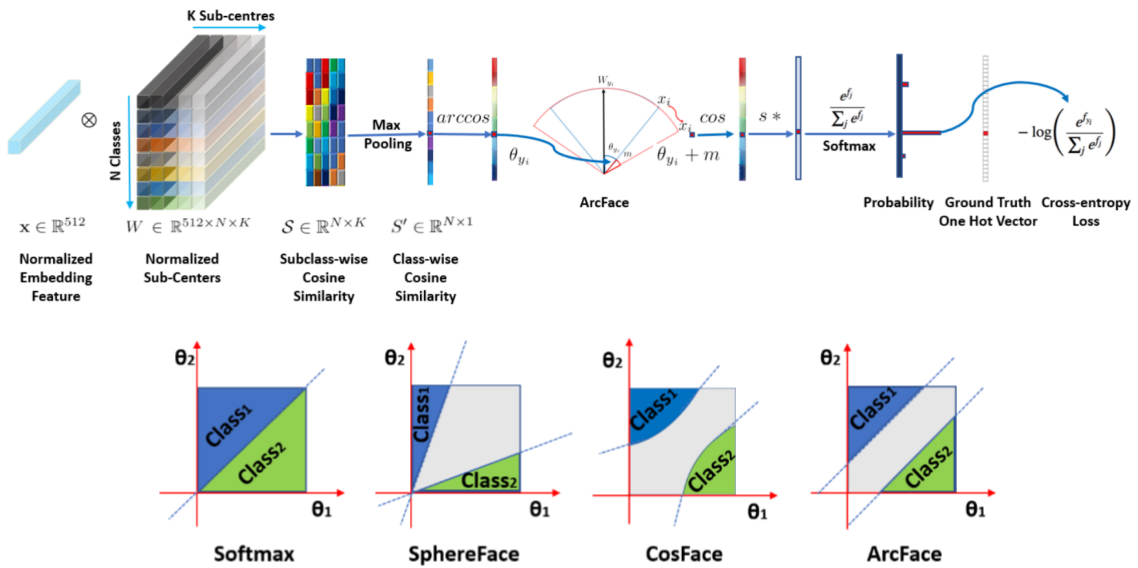


Fig. 5. Decision margins of different loss functions under binary classification case. The dashed line represents the decision boundary, and the grey areas are the decision margins.