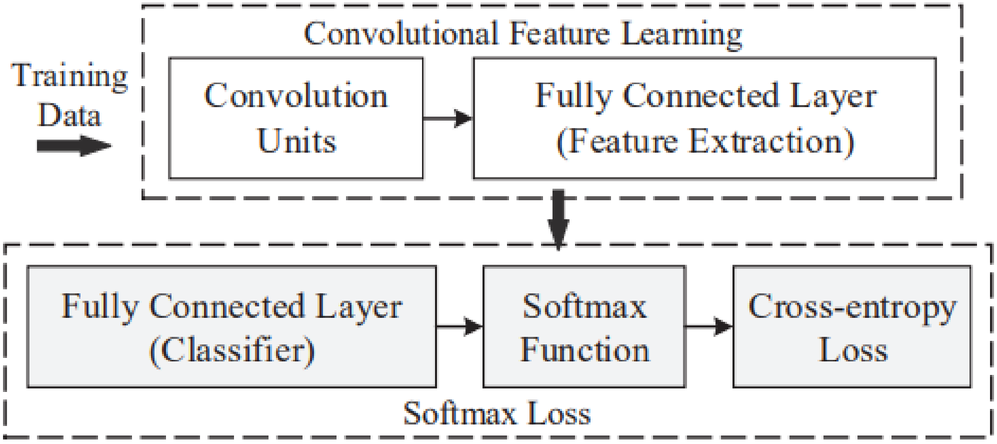
*Softmax Loss

Softmax Loss



$$
\begin{aligned}
SL &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k} * log(\frac{e^{z_{i,k}}}{\sum_{j=1}^{K} e^{z_{i,j}}}) \\
&= -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k} * log(\frac{e^{||W_{y_k}||||x_i||cos(\theta_{y_{i,k}})}}{\sum_{j=1}^{K} e^{||W_{y_{i,j}}||||x_i||cos(\theta_{y_{i,j}})}})
\end{aligned}
$$

- $y_{i,k}$表示第i个样本属于第k个类别的真是标签,当样本i属于类别k时,$y_{i,k}=1$,否则$y_{i,k}=0$
- $z_{i,k}$是样本i关于类别k的logits,**全连接的输出结果就叫做logits**
- N表示样本数
- K表示类别数
- 全连接层可以认为是基于距离函数为余弦相似度的线性分类器

L-Softmax

$$
LSL = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k} * log(\frac{e^{||W_{y_{i,k}}||||x_i||\psi(\theta_{y_{i,k}})}}{\sum_{j\neq y_{i,j}} e^{||W_{y_{i,j}}||||x_i||cos(\theta_{y_{i,j}})} + e^{||W_{y_{i,k}}||||x_i||\psi(\theta_{y_{i,k}})}})
$$

$$
\psi(\theta) = \begin{cases} cos(m\theta) & 0 \le \theta \le \frac{\pi}{m} \\ D(\theta) & \frac{\pi}{m} < \theta \le \pi \end{cases}
$$

论文中设计如下

$$
\begin{aligned}
\psi\theta &= (-1)^k cos(m\theta) - 2k \\
\theta &\in [\frac{k\pi}{m}, \frac{(k+1)\pi}{m}] \\
k &\in [0, m-1]
\end{aligned}
$$

其中 m 为正整数,m越大,分类边距越大

A-Softmax

不同于L-Softmax,将分类器权重W归一化为1

$$ASL = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k} * log(\frac{e^{||x_i||\psi(\theta_{y_{i,k}})}}{\sum_{j \neq y_{i,j}} e^{||x_i||cos(\theta_{y_{i,j}})} + e^{||x_i||\psi(\theta_{y_{i,k}})}})$$

$$\psi\theta = (-1)^k cos(m\theta) - 2k$$

$$\theta \in [\frac{k\pi}{m}, \frac{(k+1)\pi}{m}]$$

$$k \in [0, m-1]$$

AM-Softmax

Additive Angular Margin Loss

更改为了新的 $\psi(\theta) = 30 * (cos(\theta) - m)$

可以看作是对角度距离的优化,相比于余弦距离来说,当角度接近0和π时,余弦值会更密集.因此推测优化角度距离比优化余弦距离更有效果.

参考

[一文看懂softmax loss]

[理解L-Softmax、A-Softmax 和 AM-Softmax]

[AM-Softmax]

[happynear技术专栏]

[Softmax理解之margin]

[从最优化的角度看待Softmax损失函数]

[Additive Margin Softmax for Face Verification]