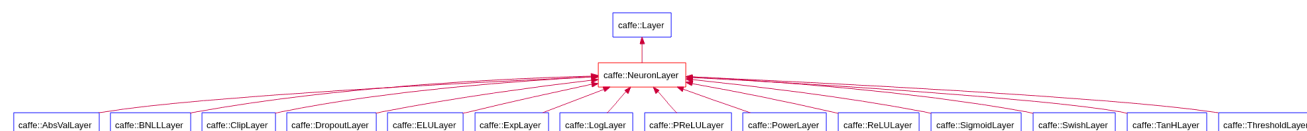


激活函数层

NeuronLayer类及其继承关系

本节主要介绍NeuronLayer类及其派生的子类关系。在caffe的include/caffe/layers/neuron_layer.hpp中写到，neuron_layer类及其派生子类的输入blob和输出blob的size都是一样的，同时输出blob的每个元素只和输入blob所对应的元素相关。



在caffe中目前派生的NeuronLayer子类有：AbsValLayer, BNLLayer, ClipLayer, DropoutLayer, ELULayer, ExpLayer, LogLayer, PReLULayer, PowerLayer, ReLULayer, SigmoidLayer, SwishLayer, TanHLayer和ThresholdLayer, 下面将会对各个派生类进行相应的介绍和测试。

NeuronLayer类

(include/caffe/layers/neuron_layer.hpp)

```
template <typename Dtype>
class NeuronLayer : public Layer<Dtype> {
public:
    explicit NeuronLayer(const LayerParameter& param)
        : Layer<Dtype>(param) {}
    virtual void Reshape(const vector<Blob<Dtype>*>& bottom,
                        const vector<Blob<Dtype>*>& top);
    virtual inline int ExactNumBottomBlobs() const { return 1; } //bottom只有一个blob
    virtual inline int ExactNumTopBlobs() const { return 1; } //top也只有一个blob
};
```

NeuronLayer重载了基类Layer的Reshape函数、ExactNumBottomBlobs函数和ExactNumTopBlobs函数。我们可以看到在NeuronLayer中bottom和top数组维度都是1，说明NeuronLayer及其派生类中是不涉及任何学习参数的，只是简单的数学映射函数的，所以我们暂称这种Layer叫做函数Layer。同时它要求经过NeruonLayer类的bottom blob和top blob的维度要一样，我们可以从其.cpp文件中可以看出来。

```
template <typename Dtype>
void NeuronLayer<Dtype>::Reshape(const vector<Blob<Dtype>*>& bottom,
                                const vector<Blob<Dtype>*>& top) {
    top[0]->ReshapeLike(*bottom[0]); //bottom blob和input blob的维度要一样
}
```

AbsValLayer类

该类比较简单，就不做具体分析。

其前向传播函数为：

$$y = |x| \quad (1)$$

反向传播函数为：

$$y' = \begin{cases} 1, & 0 \leq x \\ -1, & x < 0 \end{cases} \quad (2)$$

反向传播传到该层的的时候，会在top blob的diff中存有梯度为：

$$\frac{\partial E}{\partial y} \quad (3)$$

那传到该层的bottom blob的时候中的diff的梯度为：

$$\frac{\partial E}{\partial x} = \begin{cases} \frac{\partial E}{\partial y}, & 0 \leq x \\ -\frac{\partial E}{\partial y}, & x < 0 \end{cases} \quad (4)$$

具体的测试代码可以见(fish_test/test_absval_layer.hpp)

BNLLayer类

前向传播函数：

$$y = \begin{cases} x + \log(1 + \exp(-x)), & x > 0 \\ \log(1 + \exp(x)), & x \leq 0 \end{cases} \quad (5)$$

反向传播函数为：

$$y' = \begin{cases} \exp(50)/(1 + \exp(50)), & x > 50.0 \\ \exp(x)/(1 + \exp(x)), & x \leq 0 \end{cases} \quad (6)$$

我目前不是很清楚这个bnllayer具体是干嘛的，对于其求导也表示非常好奇。

反向传播传到该层的的时候，会在top blob的diff中存有梯度为：

$$\frac{\partial E}{\partial y} \quad (7)$$

那传到该层的bottom blob的时候中的diff的梯度为：

$$\frac{\partial E}{\partial x} = \begin{cases} \frac{\partial E}{\partial y} * \exp(50)/(1 + \exp(50)), & x > 50.0 \\ \frac{\partial E}{\partial y} * \exp(x)/(1 + \exp(x)), & x \leq 0 \end{cases} \quad (8)$$

ClipLayer类

前向传播函数为：

$$y = \max(\min, \min(\max, x)) \quad (9)$$

反向传播函数为：

$$y' = \begin{cases} 0 & \text{if } x < \min \vee x > \max \\ 1 & \text{if } x \geq \min \wedge x \leq \max \end{cases} \quad (10)$$

反向传播传到该层的的时候，会在top blob的diff中存有梯度为：

$$\frac{\partial E}{\partial y} \quad (11)$$

那传到该层的bottom blob的时候中的diff的梯度为：

$$\frac{\partial E}{\partial x} = \begin{cases} 0 & \text{if } x < \min \vee x > \max \\ \frac{\partial E}{\partial y} & \text{if } x \geq \min \wedge x \leq \max \end{cases} \quad (12)$$

DropoutLayer类

dropout是为了防止过拟合的时候使用的，但是它一般在全连接层的时候使用，而且我们只在训练的时候开启dropout，在测试的时候会将整个dropout层给关闭掉。

前向传播过程

$$y_{\text{train}} = \begin{cases} \frac{x}{1-p} & \text{if } u > p, \quad u \sim U(0, 1) \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

反向传播过程

$$y'_{\text{train}} = \begin{cases} \frac{1}{1-p} & \text{if } u > p, \quad u \sim U(0, 1) \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

反向传播传到该层的的时候，会在top blob的diff中存有梯度为：

$$\frac{\partial E}{\partial y_{\text{train}}} \quad (15)$$

那传到该层的bottom blob的时候中的diff的梯度为：

$$\frac{\partial E}{\partial x} = \begin{cases} \frac{\partial E}{\partial y_{\text{train}}} * \frac{1}{1-p} & \text{if } u > p, \quad u \sim U(0, 1) \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

注意在dropout的实现中我们需要通过伯努利分布来生成0-1数。

ELULayer类

前向传播过程

$$y = \begin{cases} x & \text{if } x > 0 \\ \alpha(\exp(x) - 1) & \text{if } x \leq 0 \end{cases} \quad (17)$$

反向传播过程

$$y' = \begin{cases} 1 & \text{if } x > 0 \\ y + \alpha & \text{if } x \leq 0 \end{cases} \quad (18)$$

反向传播传到该层的的时候，会在top blob的diff中存有梯度为：

$$\frac{\partial E}{\partial y} \quad (19)$$

那传到该层的bottom blob的时候中的diff的梯度为：

$$\frac{\partial E}{\partial x} = \begin{cases} \frac{\partial E}{\partial y} & \text{if } x > 0 \\ \frac{\partial E}{\partial y} * (y + \alpha) & \text{if } x \leq 0 \end{cases} \quad (20)$$

ExpLayer类

前向传播过程

$$y = \gamma^{\alpha x + \beta} \quad (21)$$

反向传播过程

$$y' = y \alpha \log_e(\gamma) \quad (22)$$

反向传播传到该层的的时候，会在top blob的diff中存有梯度为：

$$\frac{\partial E}{\partial y} \quad (23)$$

那传到该层的bottom blob的时候中的diff的梯度为：

$$\frac{\partial E}{\partial x} = \frac{\partial E}{\partial y} * y \alpha \log_e(\gamma) \quad (24)$$

LogLayer类

前向传播过程

$$y = \log_{\gamma}(\alpha x + \beta) \quad (25)$$

反向传播过程

$$y' = y \alpha \log_e(\gamma) \quad (26)$$

反向传播传到该层的的时候，会在top blob的diff中存有梯度为：

$$\frac{\partial E}{\partial y} \quad (27)$$

那传到该层的bottom blob的时候中的diff的梯度为：

$$\frac{\partial E}{\partial x} = \frac{\partial E}{\partial y} y \alpha \log_e(\gamma) \quad (28)$$

PReLULayer类

前向传播过程

$$y_i = \max(0, x_i) + a_i \min(0, x_i) \quad (29)$$

反向传播过程：

$$y' = \begin{cases} a_i & \text{if } x_i \leq 0 \\ 1 & \text{if } x_i > 0 \end{cases} \quad (30)$$

反向传播传到该层的的时候，会在top blob的diff中存有梯度为：

$$\frac{\partial E}{\partial y} \quad (31)$$

那传到该层的bottom blob的时候中的diff的梯度为：

$$\frac{\partial E}{\partial x_i} = \begin{cases} a_i \frac{\partial E}{\partial y_i} & \text{if } x_i \leq 0 \\ \frac{\partial E}{\partial y_i} & \text{if } x_i > 0 \end{cases} \quad (32)$$

PowerLayer类

前向传播过程：

$$y = (\alpha x + \beta)^\gamma \quad (33)$$

反向传播过程：

$$y' = \alpha \gamma (\alpha x + \beta)^{\gamma-1} = \frac{\alpha \gamma y}{\alpha x + \beta} \quad (34)$$

反向传播传到该层的的时候，会在top blob的diff中存有梯度为：

$$\frac{\partial E}{\partial y} \quad (35)$$

那传到该层的bottom blob的时候中的diff的梯度为：

$$\frac{\partial E}{\partial x} = \frac{\partial E}{\partial y} \alpha \gamma (\alpha x + \beta)^{\gamma-1} = \frac{\partial E}{\partial y} \frac{\alpha \gamma y}{\alpha x + \beta} \quad (36)$$

ReLULayer类

前向传播：

$$y = \max(0, x) \quad (37)$$

反向传播：

$$y' = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases} \quad (38)$$

反向传播传到该层的的时候，会在top blob的diff中存有梯度为：

$$\frac{\partial E}{\partial y} \quad (39)$$

那传到该层的bottom blob的时候中的diff的梯度为：

$$\frac{\partial E}{\partial x} = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{\partial E}{\partial y} & \text{if } x > 0 \end{cases} \quad (40)$$

SigmoidLayer类

前向传播：

$$y = (1 + \exp(-x))^{-1} \quad (41)$$

反向传播：

$$y' = y(1 - y) \quad (42)$$

反向传播传到该层的的时候，会在top blob的diff中存有梯度为：

$$\frac{\partial E}{\partial y} \quad (43)$$

那传到该层的bottom blob的时候中的diff的梯度为：

$$\frac{\partial E}{\partial x} = \frac{\partial E}{\partial y} y(1 - y) \quad (44)$$

SwishLayer类

前向传播：

$$y = x\sigma(\beta x) \quad (45)$$

反向传播：

$$y' = (\beta y + \sigma(\beta x)(1 - \beta y)) \quad (46)$$

反向传播传到该层的的时候，会在top blob的diff中存有梯度为：

$$\frac{\partial E}{\partial y} \quad (47)$$

那传到该层的bottom blob的时候中的diff的梯度为：

$$\frac{\partial E}{\partial x} = \frac{\partial E}{\partial y} (\beta y + \sigma(\beta x)(1 - \beta y)) \quad (48)$$

TanHLayer类

前向传播：

$$y = \frac{\exp(2x) - 1}{\exp(2x) + 1} \quad (49)$$

反向传播：

$$y' = \left(1 - \left[\frac{\exp(2x) - 1}{\exp(2x) + 1}\right]^2\right) = (1 - y^2) \quad (50)$$

反向传播传到该层的的时候，会在top blob的diff中存有梯度为：

$$\frac{\partial E}{\partial y} \quad (51)$$

那传到该层的bottom blob的时候中的diff的梯度为：

$$\frac{\partial E}{\partial x} = \frac{\partial E}{\partial y} \left(1 - \left[\frac{\exp(2x) - 1}{\exp(2x) + 1}\right]^2\right) = \frac{\partial E}{\partial y} (1 - y^2) \quad (52)$$

ThresholdLayer类

前向传播

$$y = \begin{cases} 0 & \text{if } x \leq t \\ 1 & \text{if } x > t \end{cases} \quad (53)$$

注意：并没有反向传播。