

## 1.2 行为主义、符号主义、连接主义

2.3 首先，确定一个损失函数 $L(\omega)$ ， $L(\omega)$ 应能对 $\omega$ ， $b$ 微分。求梯度的过程，实质上就是求 $\frac{\partial L(\omega)}{\partial \omega}$ 和 $\frac{\partial L(\omega)}{\partial b}$ ，利用求导的链式法则，依次求出对应的 $\frac{\partial L(\omega)}{\partial \omega}$ 和 $\frac{\partial L(\omega)}{\partial b}$ 即可。通过设定一个学习率 $\eta$ ，利用公式 $\omega \leftarrow \omega - \eta \times \frac{\partial L(\omega)}{\partial \omega}$ 来更新权重。

2.5 为了避免过拟合，可以使用多种正则化方法

- 参数范数惩罚：使用 $L^2$ 正则化或 $L^1$ 正则化，将权重的大小也添加到损失函数中，来避免权重过大的情况
- Bagging集成学习：训练不同的模型或者不同参数的同一个模型，将多个模型同时应用于决策，以发挥不同模型的决策优势
- 提前终止学习：在训练模型的过程中，当测试误差不再减低或者已经达到预定的迭代测试时，提前终止训练以避免进一步训练导致模型过拟合

2.7 对 $\Phi(v) = \frac{v}{\sqrt{1+v^2}}$ 求导数，可得

$$\Phi'(v) = \frac{\sqrt{1+v^2} - v \cdot \frac{v}{\sqrt{1+v^2}}}{1+v^2} = \frac{1}{(1+v^2)^{\frac{3}{2}}} \quad \text{代入原点 } v = 0 \text{ 得}$$

$$\Phi'(0) = 1$$

## 3.2

规定：

问题中的乘法数量和加法数量分别指的是两个矩阵之间的乘法运算和加法运算的次数

$N_{ir}$ 、 $N_{ic}$ ：分别是输入特征图的宽度和高度

$N_{if}$ ：输入特征图的个数

$N_{or}$ 、 $N_{oc}$ ：分别是输出特征图的宽度和高度

$N_{of}$ ：输出特征图的个数

$K_r$ 、 $K_c$ ：分别是卷积核的宽度和高度

$s_r$ 、 $s_c$ ：分别是水平方向和垂直方向的卷积步长

- AlexNet

经分析，AlexNet包含Conv、ReLU、LRN、MaxPool、FC、softmax共六种类型的网络层，

Conv层：乘法数量等于 $N_{if} \cdot N_{or} \cdot N_{oc} \cdot N_{of}$ ，加法数量等于 $(N_{if} - 1) \cdot N_{of}$

ReLU层：乘法数量等于加法数量等于零

LRN层：乘法数量等于加法数量等于零

MaxPool层：乘法数量等于加法数量等于零

FC层：乘法数量等于加法数量等于一

softmax层：乘法数量等于加法数量等于零

则，对于AlexNet的网络配置统计计算结果，

第一个conv层：

乘法次数—— $3 \times 55 \times 55 \times 96 = 871200$

加法次数—— $(3 - 1) \times 96 = 192$

第二个conv层：

乘法次数—— $96 \times 27 \times 27 \times 256 = 17915904$

加法次数—— $(96 - 1) \times 256 = 24320$

第三个conv层:

乘法次数—— $256 \times 13 \times 13 \times 384 = 16613376$

加法次数—— $(256 - 1) \times 384 = 97920$

第四个conv层:

乘法次数—— $384 \times 13 \times 13 \times 384 = 24920064$

加法次数—— $(384 - 1) \times 384 = 147072$

第五个conv层:

乘法次数—— $384 \times 13 \times 13 \times 256 = 16613376$

加法次数—— $(384 - 1) \times 256 = 98048$

FC1层: 乘法次数——1 加法次数——1

FC2层: 乘法次数——1 加法次数——1

FC3层: 乘法次数——1 加法次数——1

经统计,

乘法次数—— $871200 + 17915904 + 16613376 + 24920064 + 16613376 + 1 + 1 + 1 = 76933923$

加法次数—— $192 + 24320 + 97920 + 147072 + 98048 + 1 + 1 + 1 = 367555$

- VGG19

与AlexNet类似的网络组成, 只不过去掉了LRN层, 以及增加了较多卷积层。

第一个conv3-64层:

乘法次数—— $3 \times 224 \times 224 \times 64 = 9633792$

加法次数—— $(3 - 1) \times 64 = 128$

其它conv3-64层 (共1个) :

乘法次数—— $64 \times 224 \times 224 \times 64 = 205520896$

加法次数—— $(64 - 1) \times 64 = 4032$

第一个conv3-128层:

乘法次数—— $64 \times 112 \times 112 \times 128 = 102760448$

加法次数—— $(64 - 1) \times 128 = 8064$

其它conv3-128层 (共1个) :

乘法次数—— $128 \times 112 \times 112 \times 128 = 205520896$

加法次数—— $(128 - 1) \times 128 = 16256$

第一个conv3-256层:

乘法次数—— $128 \times 56 \times 56 \times 256 = 102760448$

加法次数—— $(128 - 1) \times 256 = 32512$

其它conv3-256层 (共3个) :

乘法次数—— $256 \times 56 \times 56 \times 256 = 205520896$

加法次数—— $(256 - 1) \times 256 = 65280$

第一个conv3-512层:

乘法次数—— $256 \times 28 \times 28 \times 512 = 102760448$

加法次数—— $(256 - 1) \times 512 = 130560$

第二到第四个conv3-512层 (共3个) :

乘法次数—— $512 \times 28 \times 28 \times 512 = 205520896$

加法次数—— $(512 - 1) \times 512 = 261632$

其它conv3-512层 (共4个) :

乘法次数—— $512 \times 14 \times 14 \times 512 = 51380224$

加法次数—— $(512 - 1) \times 512 = 261632$

FC层 (共3个) :

乘法次数——1      加法次数——1

经统计,

乘法次数——

$$9633792 + 205520896 \times (1 + 1 + 3 + 3) + 102760448 \times (1 + 1 + 1) + 51380224 \times 4 + 1 \times 3 = 2167603203$$

加法次数——

$$128 + 4032 + 8064 + 16256 + 32512 + 65280 \times 3 + 130560 + 261632 \times 3 + 261632 \times 4 + 1 \times 3 = 2218819$$

- ResNet152

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		$1.8 \times 10^9$	$3.6 \times 10^9$	$3.8 \times 10^9$	$7.6 \times 10^9$	$11.3 \times 10^9$

根据[原文](#)中的表1, 可以较清楚的看出ResNet152的网络结构, 其中有151个卷积层是存在矩阵乘法和加法运算的。

conv1:

乘法运算—— $3 \times 112 \times 112 \times 64 = 2408448$

加法运算—— $(3 - 1) \times 64 = 128$

conv2\_x:

乘法运算—— $56 \times 56 \times (64 \times 64 + 256 \times 64 \times 2 + 64 \times 64 \times 3 + 64 \times 256 \times 3) = 308281344$

加法运算——

$$(64 - 1) \times 64 + (256 - 1) \times 64 \times 2 + (64 - 1) \times 64 \times 3 + (64 - 1) \times 256 \times 3 = 97152$$

conv3\_x:

乘法运算——

$$28 \times 28 \times (256 \times 128 + 512 \times 128 \times 7 + 128 \times 128 \times 8 + 128 \times 512 \times 8) = 899153920$$

加法运算——

$$(256 - 1) \times 128 + (512 - 1) \times 128 \times 7 + (128 - 1) \times 128 \times 8 + (128 - 1) \times 512 \times 8 = 1140736$$

conv4\_x:

乘法运算——

$$14 \times 14 \times (512 \times 256 + 1024 \times 256 \times 35 + 256 \times 256 \times 36 + 256 \times 1024 \times 36) = 4136108032$$

加法运算——

$$(512 - 1) \times 256 + (1024 - 1) \times 256 \times 35 + (256 - 1) \times 256 \times 36 + (256 - 1) \times 1024 \times 36 = 21047296$$

conv5\_x:

乘法运算——

$$7 \times 7 \times (1024 \times 512 + 2048 \times 512 \times 2 + 512 \times 512 \times 3 + 512 \times 2048 \times 3) = 321126400$$

加法运算——

$$(1024 - 1) \times 512 + (2048 - 1) \times 512 \times 2 + (512 - 1) \times 512 \times 3 + (512 - 1) \times 2048 \times 3 = 6544384$$

经统计,

$$\text{乘法次数} = 2408448 + 308281344 + 899153920 + 4136108032 + 321126400 = 5667078144$$

$$\text{加法次数} = 128 + 97152 + 1140736 + 21047296 + 6544384 = 28829696$$