20220506-书&机器学习

1.过程描述

1.1 故事

> 文化质感仅仅是一个宇宙真理的表面表达，这便是：我们都是人类，都经历着同样根本的人类难题，提出同样根本的人类疑问，生活在不断缩减的时间阴影之下。

> 在这些人物及其冲突的深处，我们找到了自己的人性。我们去看电影，进入一个令人痴迷的新世界，去设身处地地体验一份初看起来似乎并不同于我们，而其内心却又和我们息息相通的，另一个人的生活。体验一个虚构的世界，去照亮我们的日常现实。我们并不希望逃避生活，而是希望发现生活，以焕然一新的试验性方式，去运用我们的思想，宣泄我们的情感，去欣赏，去学习，去增加生活的深度。

> 日复一日，我们寻求亚里士多德在《伦理学》中提出的那一古老问题的答案：一个人应该如何度过他的一生！但，问题的答案总是在规避着我们。当我们力图使我们的手段合乎我们的梦想时，当我们力图将我们的思想融入我们的激情时，当我们力图让我们的欲望变成现实时，那一问题的答案始终躲藏在飞速流逝、难以捕捉的时间后面。我们犹如乘坐一艘飞船，险象环生地穿行在时间隧道之中。如果

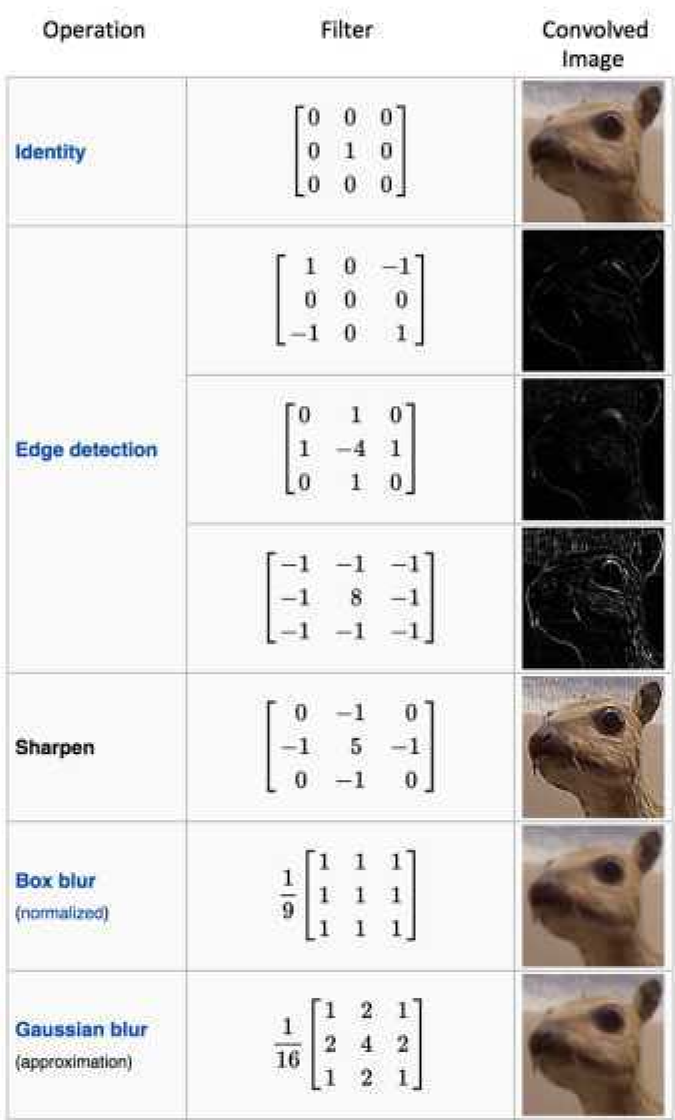
> 我们想让飞船减速，以便捕捉人生的模式和意义，人生就会像一个格式塔一样扑朔迷离：时而严肃，时而滑稽；时而静止，时而狂乱；时而意味深长，时而索然寡趣。重大的世界事务完全在我们的掌控之外，而个人事务又往往钳制着我们，尽管我们无不努力用双手牢牢掌握着自己的方向盘。

> 世人对电影、小说、戏剧和电视的消费是如此的如饥似渴、不可餍足，故事艺术已经成为人类灵感的首要来源，因为它不断寻求整治人生混乱的方法，洞察人生的真谛。我们对故事的欲望反映了人类对捕捉生活模式的深层需求，这不仅仅是一种纯粹的知识实践，而是一种非常个人化和情感化的体验。用剧作家让 阿努伊的话说：“小说赋予人生以形式。”

1.2 神经网络

今天依旧围绕着卷积神经网络进行学习，有几个印象深刻的点：

filter开展边缘检测



Operation	Filter	Convolved Image
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	
Gaussian blur (approximation)	$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	

假设图像为：

```
[[1. 1. 0. 0. 0. 0. 1. 1.]
 [1. 1. 0. 0. 0. 0. 1. 1.]
 [1. 1. 0. 0. 0. 0. 1. 1.]
 [1. 1. 0. 0. 0. 0. 1. 1.]
 [1. 1. 0. 0. 0. 0. 1. 1.]
 [1. 1. 0. 0. 0. 0. 1. 1.]]
```

filter矩阵为：

```
K = nd.array([[1, -1]])
```

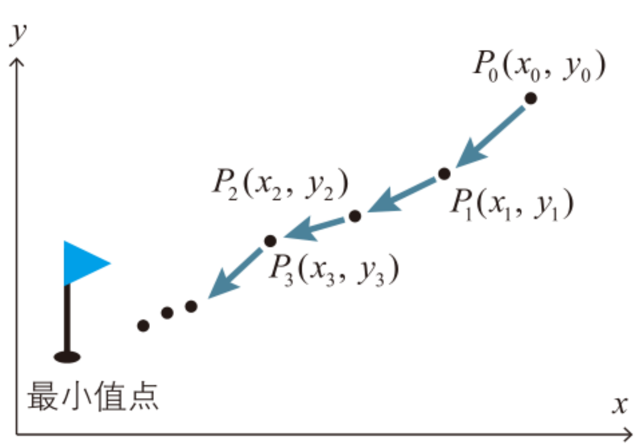
经过卷积运算后得到：

```
[[ 0.  1.  0.  0.  0. -1.  0.]
 [ 0.  1.  0.  0.  0. -1.  0.]
 [ 0.  1.  0.  0.  0. -1.  0.]
 [ 0.  1.  0.  0.  0. -1.  0.]
 [ 0.  1.  0.  0.  0. -1.  0.]
 [ 0.  1.  0.  0.  0. -1.  0.]]
```

可以看到，边缘被检测出来了。

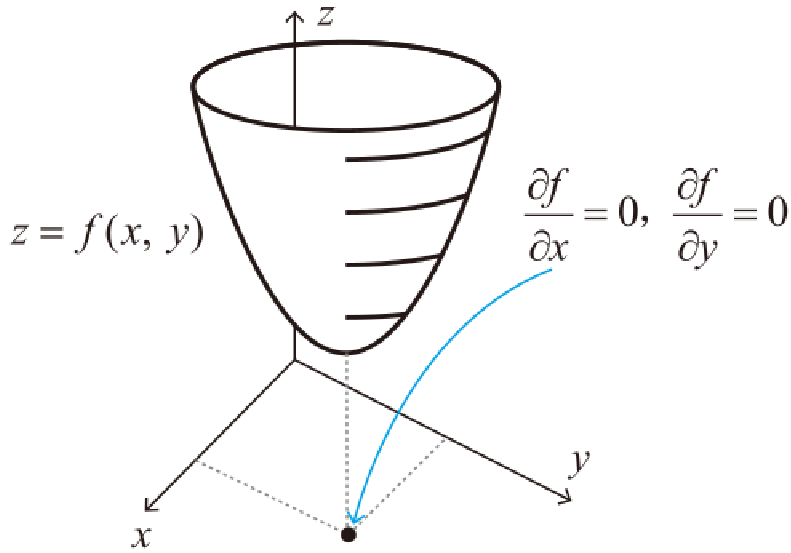
梯度下降法

梯度下降可以简单理解为，点沿着下降最快的方向移动一定的距离，在到达新的位置再寻找新的下降最快的方向并移动一定距离，通过不断重复上述过程，最终使得点达到最小值处。



从初始位置 P_0 出发，利用式 (5)、(6) 求出最陡坡度的点 P_1，然后从 P_1 出发，利用式 (5)、(6) 进一步求出最陡坡度的点 P_2，即反复利用式 (5)、(6)，最终得以最快速地到达最小值点。这就是梯度下降法。

偏导为0是函数取得极值的必要条件：



根据近似公式:

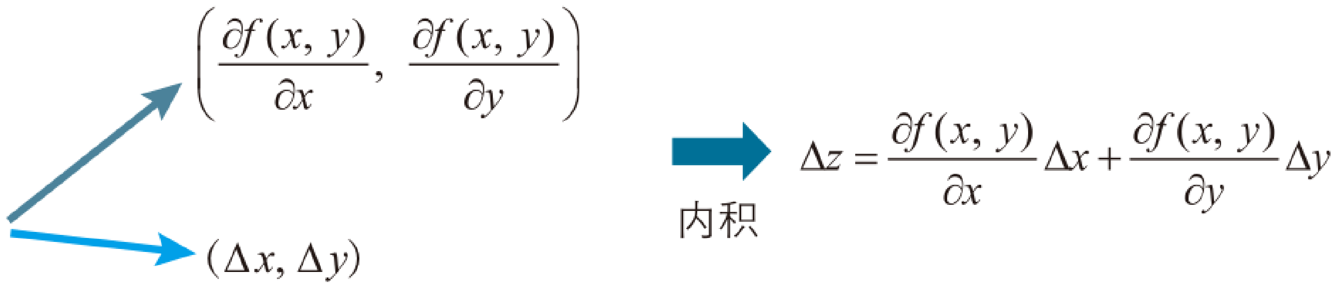
$$f(x+\Delta x) \fallingdotseq f(x) + f'(x)\Delta x \quad （\Delta x \text{为微小的数}）$$

$$f(x+\Delta x,\ y+\Delta y) \fallingdotseq f(x,\ y) + \frac{\partial f(x,\ y)}{\partial x}\Delta x + \frac{\partial f(x,\ y)}{\partial y}\Delta y$$

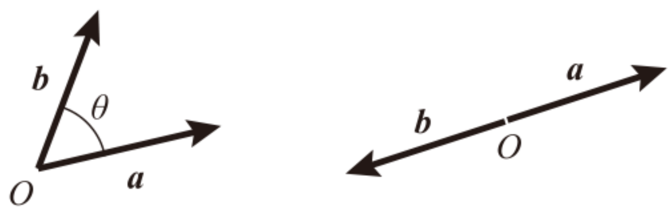
$$\Delta z = f(x+\Delta x,\ y+\Delta y) - f(x, y)$$

$$\Delta z = \frac{\partial f(x,\ y)}{\partial x}\Delta x + \frac{\partial f(x,\ y)}{\partial y}\Delta y$$

可以表示为向量内积的形式:



而当两个向量方向相反时，内积取得最小值：



向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 的内积为 $|\boldsymbol{a}||\boldsymbol{b}|\cos\theta$（$\theta$ 为两个向量的夹角）（左图）。θ 为 180° 时（即 $\boldsymbol{a}$、$\boldsymbol{b}$ 方向相反），内积的值最小（右图）。

换句话说，当向量 $\boldsymbol{b}$ 满足以下条件式时，可以使得内积 $\boldsymbol{a}\cdot\boldsymbol{b}$ 取最小值。

$$b = -ka \quad （k 为正的常数） \tag{4}$$

由此得到梯度下降基本公式：

$$(\Delta x, \Delta y) = -\eta\left(\frac{\partial f(x,y)}{\partial x}, \frac{\partial f(x,y)}{\partial y}\right) \quad （\eta 为正的微小常数）$$

其中，右边括号前面的符号即为学习率（也可称为步长，但严格来说，右边一整串才是真正移动的距离。可以对括号内的值先进行归一化，这样才便能确保步长一致。学习率需要通过不断试验进行确定。

一二阶微分用于边缘检测

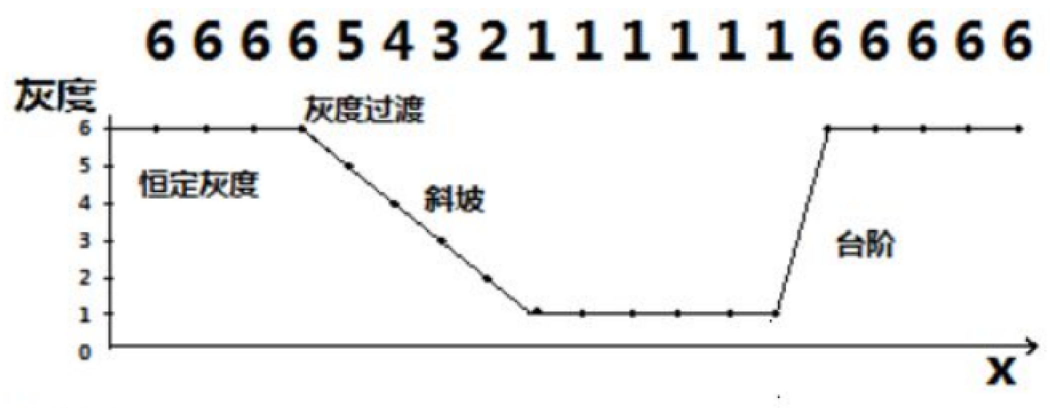
一阶微分：

$$\frac{\partial f}{\partial x} = f(x+1) - f(x)$$

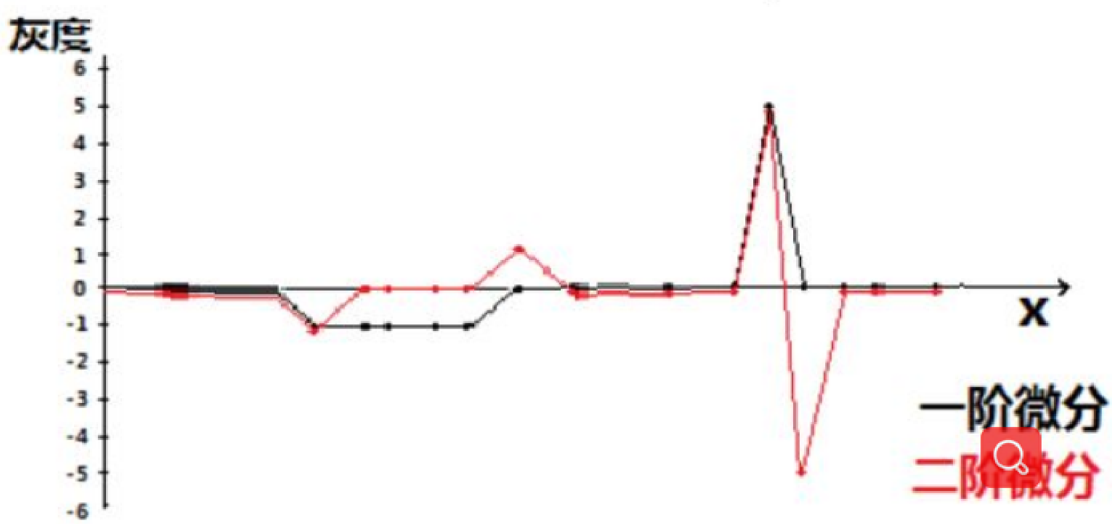
二阶微分：

$$\frac{\partial^2 f}{\partial x^2} = f(x+1) + f(x-1) - 2f(x)$$

假设有如下的灰度图数据：



应用一二阶微分后得到：



可以看到，再边缘处，二阶微分值非常大，而在其它地方值则比较小或者接近0。说明微分算子（尤其是二阶微分）对边缘图像非常敏感。

一个二维图像的拉普拉斯算子定义为：

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

$$\frac{\partial^2 f}{\partial x^2} = f(x+1,y)+f(x-1,y)-2f(x,y)$$

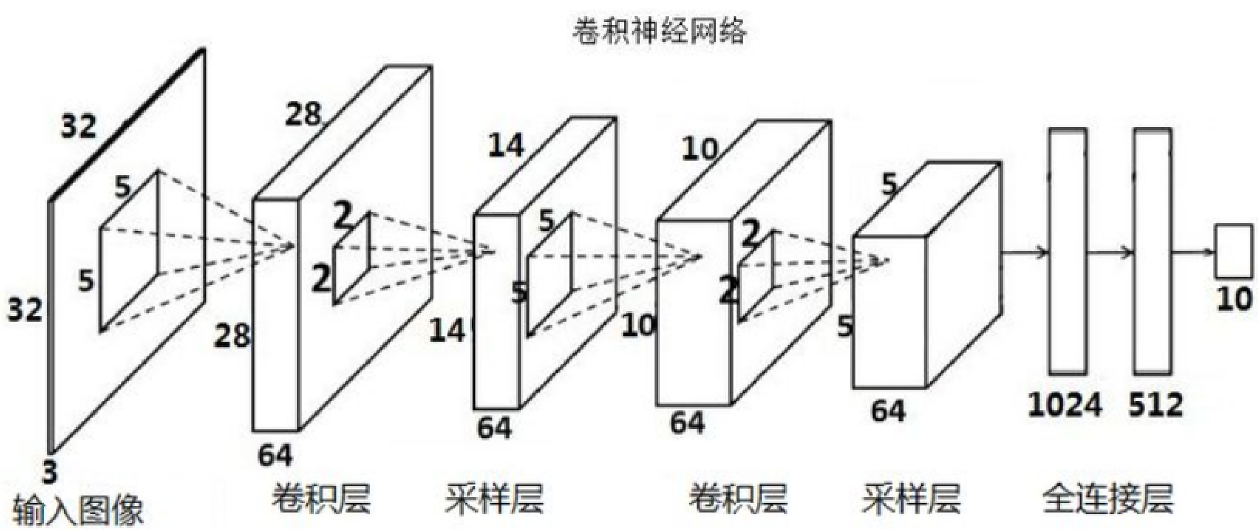
$$\frac{\partial^2 f}{\partial y^2} = f(x,y+1)+f(x,y-1)-2f(x,y)$$

$$\therefore \nabla^2 f = \frac{\partial^2 f}{\partial x^2}+\frac{\partial^2 f}{\partial y^2}$$

$$= f(x+1,y)+f(x-1,y)+f(x,y+1)+f(x,y-1)-4f(x,y)$$

0	1	0
1	-4	1
0	1	0

卷积神经网络的简单结构



这里的采样层就是之前提到的池化层。卷积神经网络大致的工作机制可以理解为：

1）通过第一个卷积层提取最初特征，输出特征图；

2）通过第一个采样层对最初的特征图进行特征选择，去除多余特征，重构新的特征图；

3）第二个卷积层是对上一层的采样层的输出特征图进行二次特征提取；

4）第二个采样层也对上层输出进行二次特征选择；

4）全连接层就是根据得到的特征进行分类。

采样层/池化层的作用

采样层其实也可以理解用用filter进行卷积运算，只是这个filter的函数机制与一般卷积层用的filter不一样。采样层可以一定程度提高空间不变性。

特征提取的误差主要来自两个方面：

1）领域大小受限；

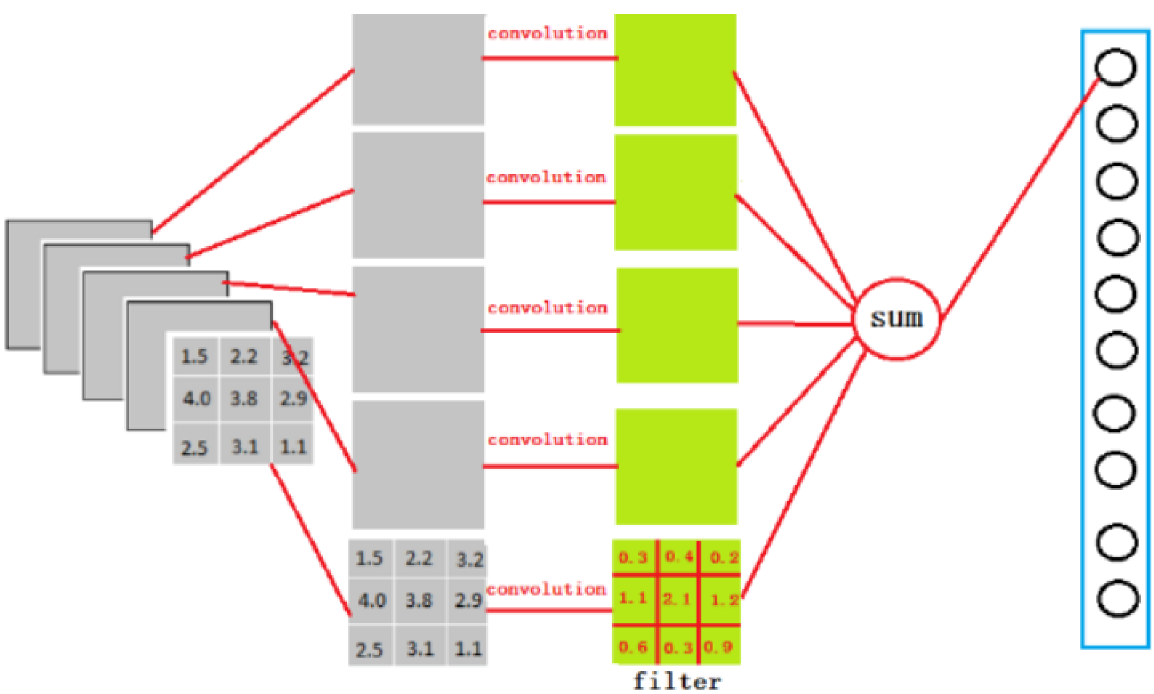
2）卷积层权值参数误差。

而average-pooling和max-pooling的主要区别在于：

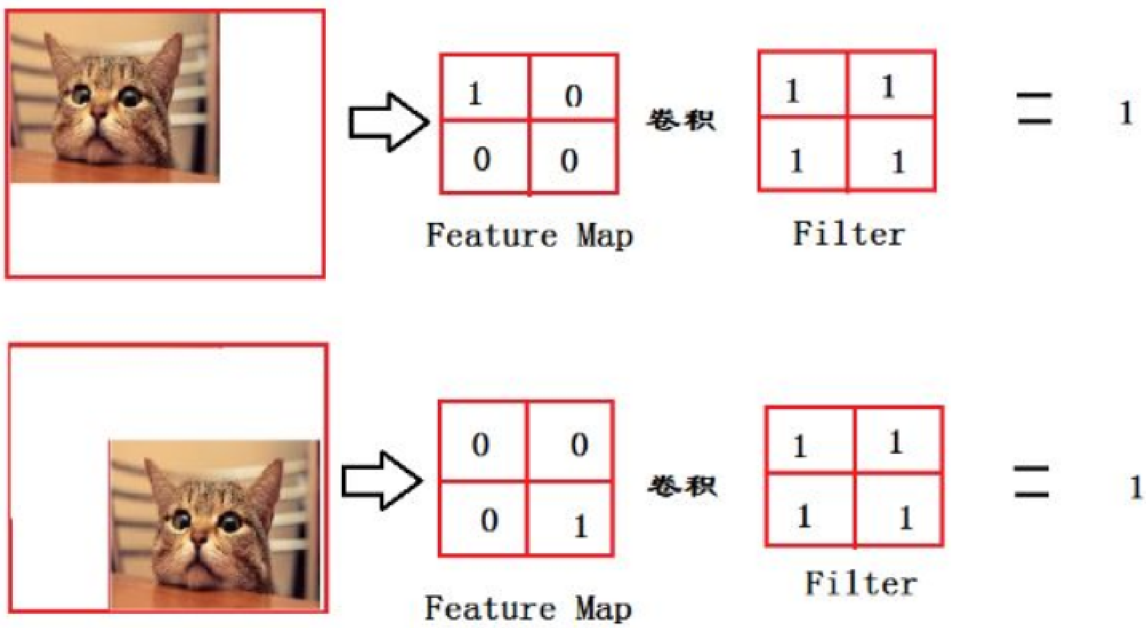
1）average-pooling能减小第一种乌哈，更多的保留图像的背景信息；

2）max-pooling能减小第二种误差，更多的保留纹理信息。

全连接层



全连接层的作用主要是实现分类。全连接层中每个神经元节点的获得同样可以视为对前一层的输出做一次卷积运算。这一步卷积的重要作用，就是把分布式特征representation映射到样本标记空间（把特征representation整合到一起，输出为一个值）。这么做的好处是大大减少特征位置对分类带来的影响。

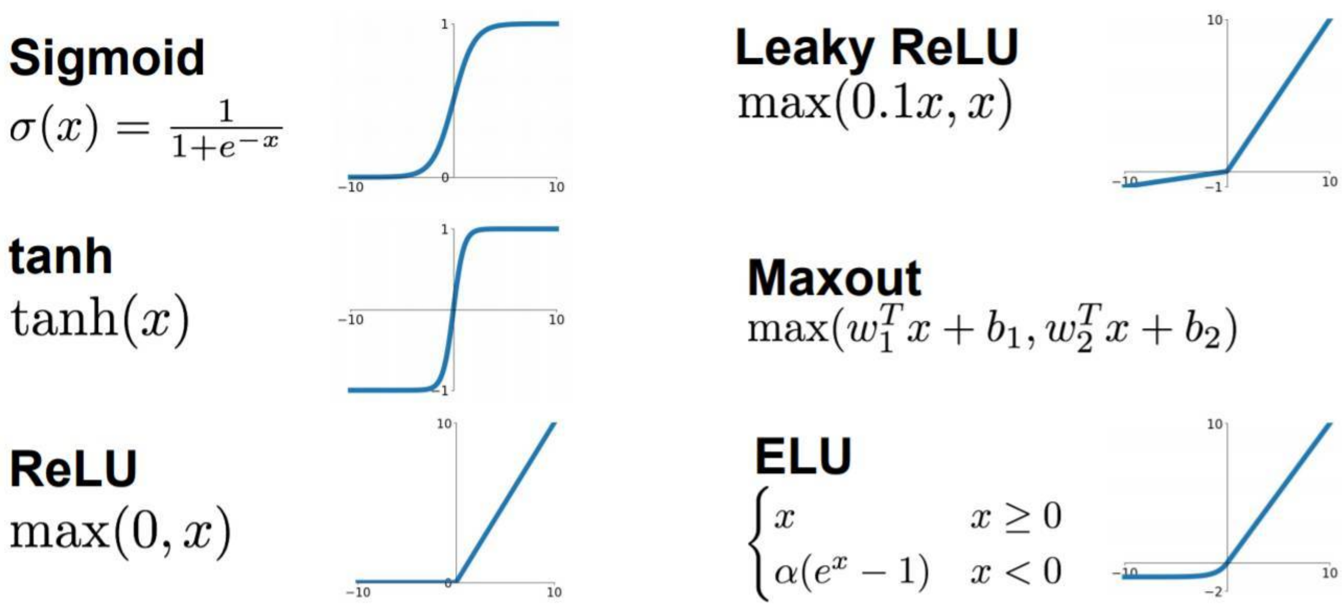


不管猫在哪里，都能用这个filter把猫找到。

此外，之所以要有两层甚至多层全连接层，是因为这样可以比较好地解决非线性问题。

全连接层的参数特别多，一些性能优异的网络如ResNet和GoogleLeNet等均用全局平均池化（GAP）取代全连接层来融合学到的深度特征。

激活函数



激活函数对于提高模型鲁棒性、**非线性表达能力**、缓解梯度消失问题、将特征图映射到新的特哼空间从而更有利于训练、加速模型收敛等问题都有很好的帮助

2.结果输出

今天下午稍微看了一点罗伯特麦基写的《故事》一书，去年差不多这个时候其实已经看了一部分，但当时没坚持多久就放弃了。现在重新翻看，突然发现里面很多话都挺有启示性的。这本书主要讲的是如何写好、呈现好一个电影故事，之所以读它也是出于比较功利性的目的：想成为一个还可以的故事讲述者。一直觉得自己在做汇报或者讲演时表现得不够好，因此便想着通过学习一些原理性、结构性的东西来构建一个普遍适用的框架，以更从容不迫地应对此类情景。此外，一直以来也觉得自己似乎在导演方面有一点小小的天赋（这种感觉主要来自于自己跟同学合作排演一些课堂小节目的经历。印象中自己在这个过程中十分活跃及主动。可惜实践的机会太少，估计也是自己的一种错觉吧），如果通过阅读这本书能让这种天赋得以激发，那岂不是美事一件。

白天的大部分时间主要在看卷积神经网络相关的书或文章。此前一直有把这块知识彻底搞懂的动力，也有好几回尝试着看一些相关的入门教程，但总遇到这样那样的学习障碍，之后就不了了之了。昨天无意中看到涌井良幸写的《深度学习的数学》，突然发现写的十分通俗易懂，对基本的CNN结构也有了比较清晰的认知，虽然内心依旧有这样那样的疑问，也没开始用代码实现，但感觉比起之前而言认知上算是跃升了一大步。明天应该还是围绕着这个主题进行学习。