

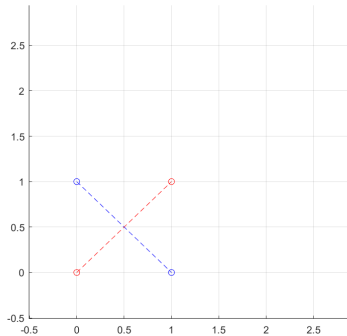
2.早期的单层感知器可认为是最简单的人工神经网络，它只能学习线性可分模式。1969 年 Minsky 和 Papert 在名为 Perceptrons 一书中表明，这类网络不能学习 XOR 函数。

a) 前馈神经网络中，常用的激活函数有 Sigmoid, ReLU 等。如果不使用这些非线性激活函数，请问会产生什么问题。

若不使用非线性层，多个线性全连接层的叠加依然相当于一个线性层：

$$y = W_n x_{n-1} + b_n = W_n((W_{n-1}(W_{n-2} \dots) + b_{n-1}) + b_n = W_n W_{n-1} \dots W_1 x + b'$$

因此只能学习线性可分模式，表现力差。一个例子是不能学到 XOR 函数。下图可见由于直线 $(0, 0) - (1, 1)$ 与 $(0, 1) - (1, 0)$ 相交，不可能用直线将两种输入加以区分。



b) 试设计一个前馈神经网络来解决 XOR 问题，要求该前馈网络具有两个隐藏神经元和一个输出神经元，要求激活函数为 RELU。请给出该神经网络的具体参数值，并验证该组参数值能满足要求。

分析：XOR 是二输入函数，输入两个 0-1 值 x_1, x_2 ，输出它们异或后的 0-1 值：

$$XOR(x_1, x_2) : (0, 0) \rightarrow 0, (1, 1) \rightarrow 0, (1, 0) \rightarrow 1, (0, 1) \rightarrow 1$$

设计："异或"代表两个输入 x_1, x_2 不相同，自然首先想到通过线性层 1 将输入作差，得到 $x_1 - x_2$ 与 $x_2 - x_1$ 。而随后又容易想到将该差通过 ReLU；一方面是为了防止 $x_1 - x_2$ 和 $x_2 - x_1$ 相互抵消；另一方面则是反映"只要感知到一个输入比另外一个输入大 1 就激活"。最后只需将 ReLU 的输出通过线性层 2 进行求和（或者说取"或"）。因此设计网络如下：

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \xrightarrow{L1} \begin{pmatrix} x_1 - x_2 \\ x_2 - x_1 \end{pmatrix} \xrightarrow{ReLU} \begin{pmatrix} \max(x_1 - x_2, 0) \\ \max(x_2 - x_1, 0) \end{pmatrix} \xrightarrow{L2} \max(x_1 - x_2, 0) + \max(x_2 - x_1, 0)$$

输出 $\max(x_1 - x_2, 0) + \max(x_2 - x_1, 0)$ 等同于 $XOR(x_1, x_2)$ 。网络中线性层 1 ($L1$) 的参数为：

$$W_1 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, b_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

线性层 2 ($L2$) 的参数为：

$$W_2 = (1 \quad 1), b_2 = 0$$

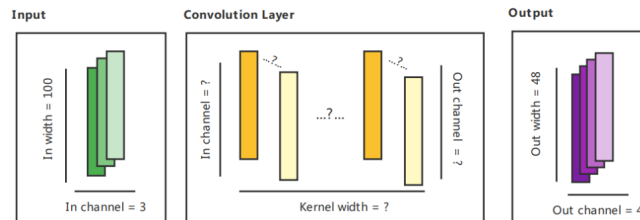
验证：将 (x_1, x_2) 的四种取值代入，发现满足。

3.含卷积层神经网络的计算。

a) 给定某卷积神经网络如下图所示，含输入层，卷积层与输出层。给定卷积步长（stride）为 2，填充宽度（padding）为 0，以及特征输入宽度为 100，输入通道为 3，特征输出宽度为 48，输出通道为 4，请计算卷积层输入输出通道的值，及卷积核宽度的值。

卷积层输入通道为3，输出通道为4。

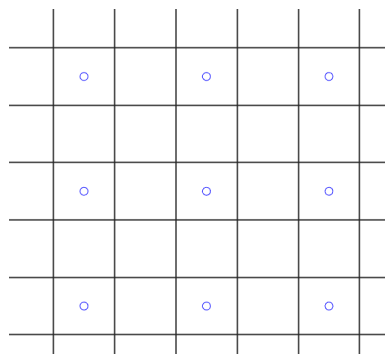
卷积核宽度可能是5或者6。由于输出宽度为48，步长为2，可知卷积起点依次为1, 3, 5, ..., 95。而由于输入宽度为100，填充宽度为0，卷积核宽度范围是(4, 6]，可能是5或者6。（一般是5，因为奇数卷积核可以将卷积结果和中心点对应）



b) 对一个步幅为 2×2 ，核大小为 3×3 的卷积层给定输入，忽略边界，为了能够在输出的特征映射中保留输入的全部信息，试讨论卷积核个数的最小值。

四个卷积核。"在输出的特征映射中保留输入的全部信息"，即可由卷积结果唯一确定输入的所有值。

由于步长为 2×2 ，卷积核为 3×3 ，可以如下图，作出题目中卷积的过程。图中一个方格对应输入的一个元素，一个蓝点对应一次卷积的中心，也即一个卷积结果或一个方程。可以看出卷积结果和输入变量的数量之比为1：4。



那么为了能还原输入信息，方程数（卷积结果数）要和未知数（输入变量数）要相等。方程数量至少要 $\times 4$ ，因此至少需要四个卷积核。

（同时这四个卷积核应线性无关）