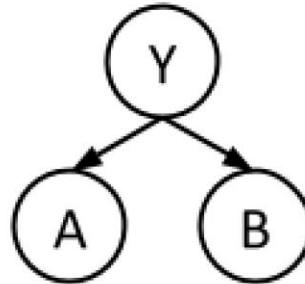


## 习题五：机器学习（共 75 分）

### 1、朴素贝叶斯（15 分）

在这个问题中，我们将训练一个朴素贝叶斯类来预测类标签  $Y$  作为输入特征的函数  $A$  和  $B$ 。 $Y$ 、 $A$  和  $B$  都是二进制变量，域为 0 和 1。我们有 10 条训练数据，用来估计我们的分布。我们的数据和模型如下图所示：

$A$	1	1	1	1	0	1	0	1	1	1
$B$	1	0	0	1	1	1	1	0	1	1
$Y$	1	1	0	0	0	1	1	0	0	0



(a)  $P(Y)$ 、 $P(A|Y)$  和  $P(B|Y)$  的最大似然估计是多少？（5 分）

$Y$	$P(Y)$	$A$	$Y$	$P(A Y)$	$B$	$Y$	$P(B Y)$
0		0	0		0	0	
1		1	0		1	0	
		0	1		0	1	
		1	1		1	1	

(b) 考虑一个新的数据点 ( $A = 1$ ,  $B = 1$ )。这个分类器将为此样本分配什么标签？（5 分）

(c) 让我们使用拉普拉斯平滑来平滑我们的分布。假如使用  $k = 2$  的拉普拉斯平滑，计算  $P(A|Y)$  的新分布。（5 分）

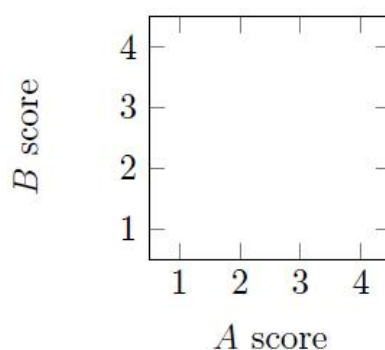
$A$	$Y$	$P(A Y)$
0	0	
1	0	
0	1	
1	1	

## 2、感知器（20 分）

您想根据电影的剧本预测电影是否会赢利。您聘请了两个评论家 A 和 B 来阅读您拥有的剧本，并以 1 到 4 的等级对其进行评分。批评者并不完美；以下是五个数据点，包括影评人的评分和电影的表现：

#	Movie Name	A	B	Profit?
1	Pellet Power	1	1	-
2	Ghosts!	3	2	+
3	Pac is Bac	2	4	+
4	Not a Pizza	3	4	+
5	Endless Maze	2	3	-

(a) 首先，您要检查数据的线性可分离性。在下面的 2D 平面上绘制数据；用 + 标记赢利的电影，用 - 标记不赢利的电影，并确定数据是否线性可分离。（5 分）



(b) 现在，您决定使用感知器对数据进行分类。假设您直接使用上面给出的分数作为特征，并使用一个偏置特征（bias）。即  $f_0 = 1$ ,  $f_1 = A$  给出的分数,  $f_2 = B$  给出的分数。使用感知器算法对数据进行一次遍历，将结果填入下表。按数据点的顺序，例如在步骤 1 中使用数据点 1，以此类推。（5 分）

step	Weights	Score	Correct?
1	$[-1, 0, 0]$	$-1 \cdot 1 + 0 \cdot 1 + 0 \cdot 1 = -1$	yes
2			
3			
4			
5			

最终的权重（Weights）：

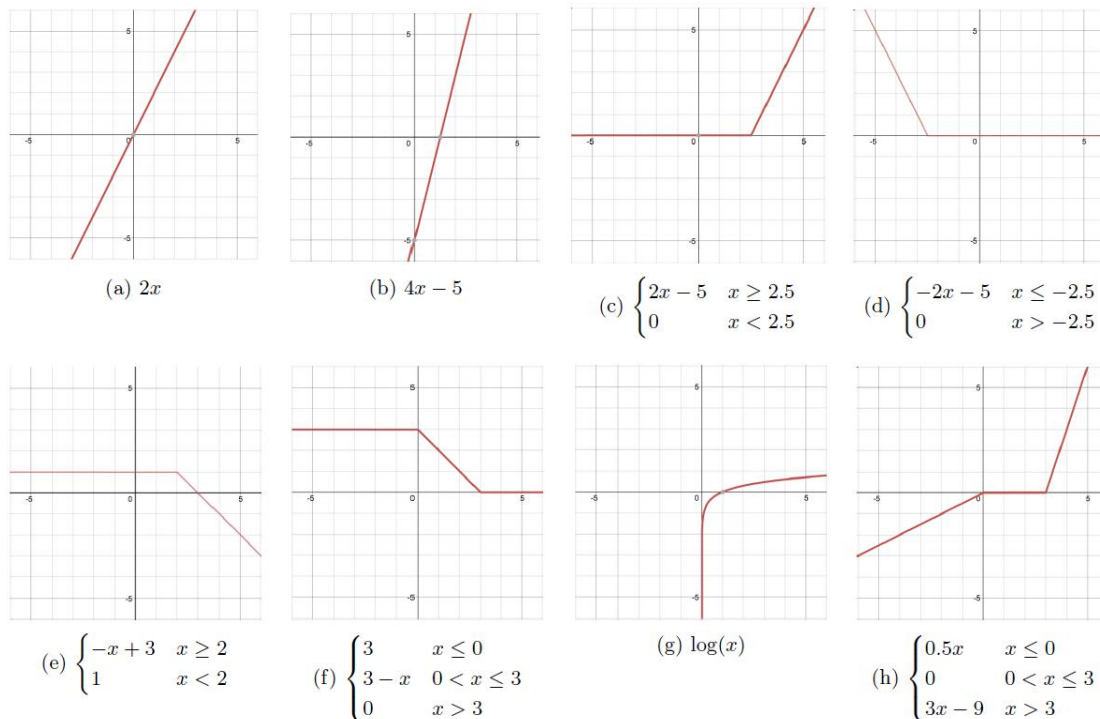
(c) 你的算法是否学会了分离数据的权重（weights）？（4 分）

(d) 更一般地说，无论训练数据如何，您都想知道您的特征是否足够强大，能够允许你处理各种情形。圈出以下场景中，使用以上特征的感知器确实可以对能否赢利的电影作出分类的例子：（6 分）

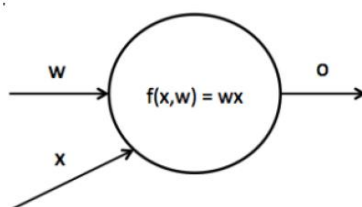
- 1) 你的评论者很棒：如果他们的总分超过 8 分，那么电影会赢利，否则就不会。
- 2) 您的评论家是艺术评论家：电影将赢利，当且仅当每个评论者给 2 或者 3 分。
- 3) 你的审稿人品味奇特且迥异。您的电影将赢利当且仅当两者的评分相同。

### 3、神经网络（40 分）

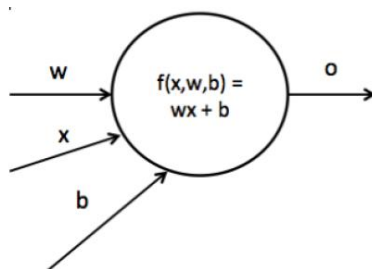
考虑单变量  $x$  的多个函数（a-h），如下图所示。后续页面上的神经元图开始简单，但会越来越复杂，逐渐建立起完整的神经网络。对每种图，指出它们能够表示以下的哪些函数。



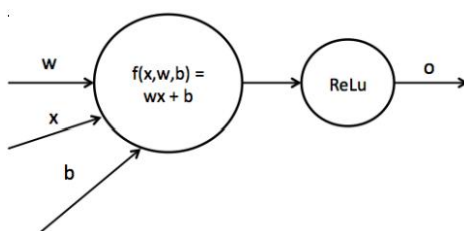
- (a) 考虑下图，计算一个线性变换，采用标量输入  $x$ 、权重  $w$ ，输出  $o$ ，使得  $o = wx$ 。这个变换可以表示上图（a-h）种的哪些函数？对于可以表达的函数，写出适当值的  $w$  值。（5 分）



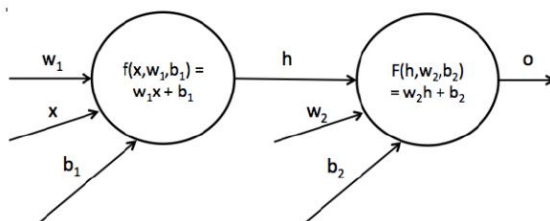
- (b) 现在我们在图中引入一个偏置项  $b$ ，使得  $o = wx + b$ （这被称为仿射函数）。这个变换可以代表上图（a-h）中的哪些函数？对于可以表达的函数，请给出适当的  $w$  和  $b$  值。（5 分）



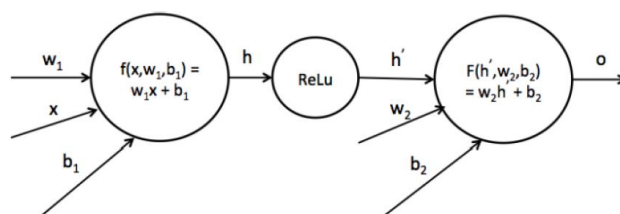
- (c) 我们可以引入非线性，如下所示。我们使用 ReLU 函数， $ReLU(x) = \max(0; x)$ 。现在上图 (a-h) 中的哪些函数可以由这个神经网络来表示？对于可以的，给出适当的  $w, b$  值。（5 分）



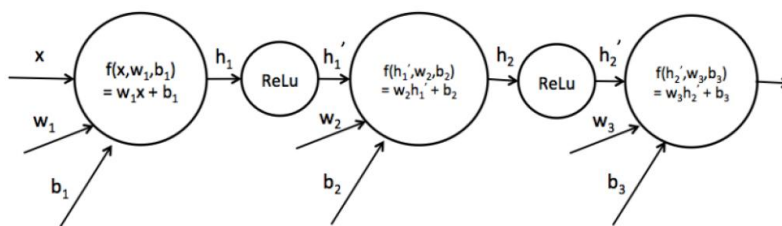
- (d) 现在我们考虑具有多个仿射变换的神经网络，如下所示。我们现在有两组权重和偏差  $w_1, b_1$  和  $w_2, b_2$ 。我们有第一层变换的结果  $h = w_1x + b_1$ ，和最终结果  $o = w_2h + b_2$ 。这个网络可以表达 (a-h) 中的哪些函数？对于可以表达的函数，请写出适当的  $w_1, b_1$  和  $w_2, b_2$  值。（5 分）



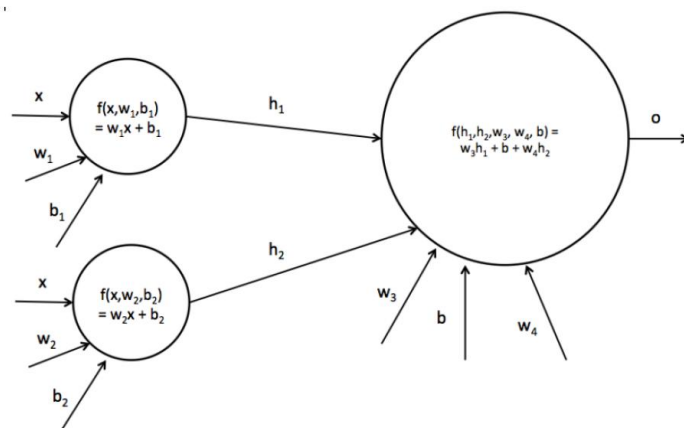
- (e) 接下来，我们在第一个仿射变换后向网络添加一个 ReLU 非线性，创建一个隐藏层。这个网络可以代表哪些函数？对于可以表达的函数，请写出适当的  $w_1, b_1$  和  $w_2, b_2$  值。（5 分）



- (f) 现在我们向网络添加另一个隐藏层，如下所示。哪些函数可以是以这个网络来表达？（5 分）



- (g) 我们想考虑使用只有一个隐藏层的神经网络，但让它更大——隐藏层的尺寸为 2。让我们考虑只使用两个仿射函数，两者之间没有非线性。这个网络可以代表那些函数？（5 分）



- (h) 现在我们在两个仿射层之间添加一个非线性，产生下面的神经网络，其中包含尺寸为 2 的隐藏层。这个网络可以代表哪些函数？（5 分）

