Homework 03 Neural Network

By Yubei Xiao

(a)

由梯度下降公式可知，$w_{1,2}^{[1]}$的梯度下降式子为：

$w_{1,2}^{[1]} = w_{1,2}^{[1]} - \alpha\frac{\partial l}{\partial w_{1,2}^{[1]}}$

而$\because \frac{\partial l}{\partial w_{1,2}^{[1]}} = \frac{\partial l}{\partial o} * \frac{\partial o}{\partial h_2} * \frac{\partial h_2}{\partial w_{1,2}^{[1]}}$，其中$o, h_2$,分别是output neuron的输出以及第二个hidden neuron的输出。

又$\because l = \frac{1}{m}\sum_{i=1}^{m}(o^{(i)} - y^{(i)})^2$

$\therefore \frac{\partial l}{\partial o} = \frac{1}{m}\sum_{i=1}^{m} 2 * (o^{(i)} - y^{(i)})$，其中上标$^{(i)}$表示的是第$i$个样本的对应的output $o^{(i)}$或者ground truth $y^{(i)}$，下面的式子中上标$^{(i)}$表达类似。

又$\because o^{(i)} = f(w_0^{[2]} + w_1^{[2]}h_1^{(i)} + w_2^{[2]}h_2^{(i)} + w_3^{[2]}h_3^{(i)}) = f((w^{[2]})^T h^{(i)})$，其中$f$为sigmoid function，而$h_j^{(i)}$对应的是第$i$个样本的第$j$个hidden neuron的输出。

$\therefore \frac{\partial o}{\partial h_2} = w_2^{[2]} f((w^{[2]})^T h^{(i)})(1 - f((w^{[2]})^T h^{(i)}))$

而$\because h_2^{(i)} = f(w_{0,2}^{[1]} + w_{1,2}^{[1]}x_1^{(i)} + w_{2,2}^{[1]}x_2^{(i)}) = f((w_2^{[1]})^T x^{(i)})$，其中$w_2^{[1]}$为从$x^{(i)}$到第二个hidden neuron的输入权重向量。

$\therefore \frac{\partial h_2}{\partial w_{1,2}^{[1]}} = x_1^{(i)} f((w_2^{[1]})^T x^{(i)})(1 - f((w_2^{[1]})^T x^{(i)}))$

因此，$w_{1,2}^{[1]}$的梯度下降式子为：

$w_{1,2}^{[1]} = w_{1,2}^{[1]} - \alpha\frac{\partial l}{\partial w_{1,2}^{[1]}}$

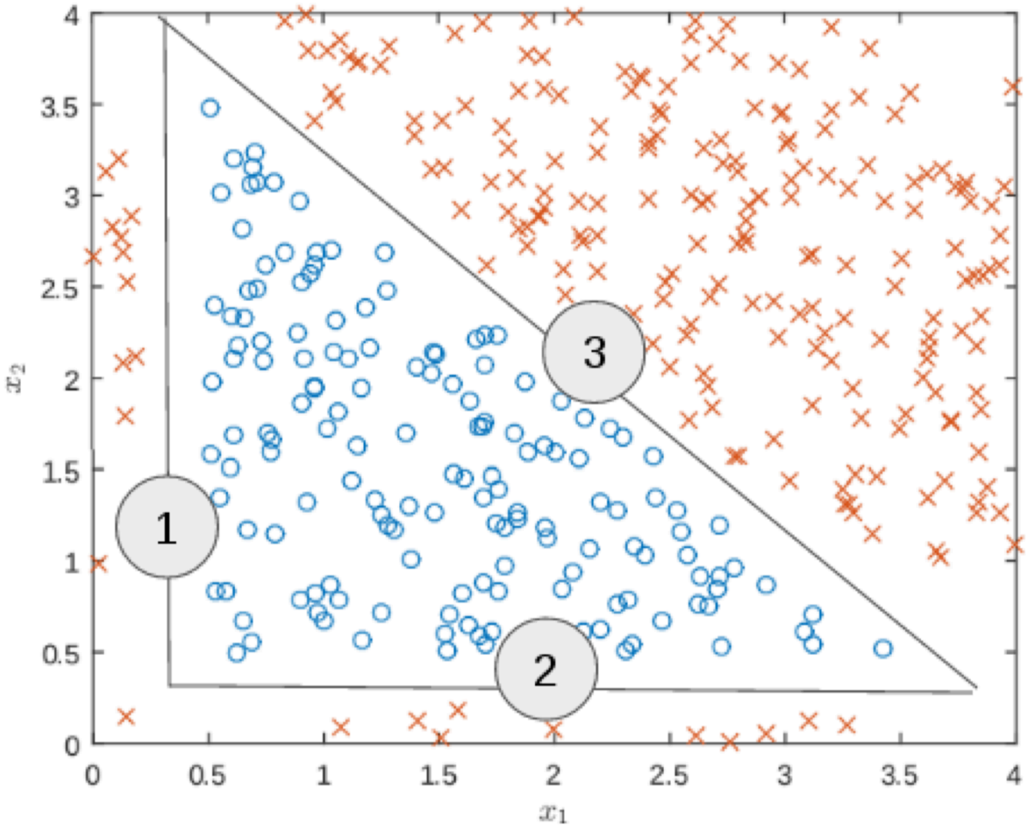
$= w_{1,2}^{[1]} - \alpha\frac{\partial l}{\partial o} * \frac{\partial o}{\partial h_2} * \frac{\partial h_2}{\partial w_{1,2}^{[1]}}$

$= w_{1,2}^{[1]} - \alpha * \frac{1}{m}\sum_{i=1}^{m} 2 * (o^{(i)} - y^{(i)}) * w_2^{[2]} f((w^{[2]})^T h^{(i)})(1 - f((w^{[2]})^T h^{(i)})) * x_1^{(i)} f((w_2^{[1]})^T x^{(i)})(1 - f((w_2^{[1]})^T x^{(i)}))$

$= w_{1,2}^{[1]} - \alpha\frac{2}{m}\sum_{i=1}^{m}(o^{(i)} - y^{(i)}) w_2^{[2]} f((w^{[2]})^T h^{(i)})(1 - f((w^{[2]})^T h^{(i)})) x_1^{(i)} f((w_2^{[1]})^T x^{(i)})(1 - f((w_2^{[1]})^T x^{(i)}))$

(b)

由提示可知，其实是找出三条线构成三角形，将两个类别分隔开来（如下图），这三条线便是由该神经网络的作用。



那么线1公式为：$-x_1 + 0.4 = 0$

线2公式为：$-x_2 + 0.3 = 0$

线3公式为：$x_1 + x_2 - 4.5 = 0$

这三条线不唯一，只需要能够将两个类别样本完全分开即可。

对于从输入x到hidden neuron的神经网络方程为：

$$f(\begin{bmatrix} w_{0,1}^{[1]} & w_{1,1}^{[1]} & w_{2,1}^{[1]} \\ w_{0,2}^{[1]} & w_{1,2}^{[1]} & w_{2,2}^{[1]} \\ w_{0,3}^{[1]} & w_{1,3}^{[1]} & w_{2,3}^{[1]} \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}) = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$$

那么对应的我们用神经网络的参数来表示出这三条线：

$$\begin{bmatrix} w_{0,1}^{[1]} & w_{1,1}^{[1]} & w_{2,1}^{[1]} \\ w_{0,2}^{[1]} & w_{1,2}^{[1]} & w_{2,2}^{[1]} \\ w_{0,3}^{[1]} & w_{1,3}^{[1]} & w_{2,3}^{[1]} \end{bmatrix} = \begin{bmatrix} 0.4 & -1 & 0 \\ 0.3 & 0 & -1 \\ -4.5 & 1 & 1 \end{bmatrix}$$

那么任意取class为0的样本，即蓝色圈样本，带入三个公式都可以得到左式< 0，那么即step function内的值< 0，而因为对于step function有：

$$f(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$$

所以h_1, h_2, h_3均$= 0$。

同理，对于class为1的样本，即红色叉样本，带入三个公式，至少可以得到一个左式> 0，相应的h_1, h_2, h_3至少有一个值$= 1$(其他为0)。

对于从hidden neuron输出h_1, h_2, h_3到output neuron的神经网络方程为：

$$f(\begin{bmatrix} w_0^{[2]} & w_1^{[2]} & w_2^{[2]} & w_3^{[2]} \end{bmatrix} \begin{bmatrix} 1 \\ h_1 \\ h_2 \\ h_3 \end{bmatrix}) = o$$

其中f也是step function，因此参数可以取：

$$\begin{bmatrix} w_0^{[2]} & w_1^{[2]} & w_2^{[2]} & w_3^{[2]} \end{bmatrix} = [-1 \quad 1 \quad 1 \quad 1]$$

则当任意class为0的样本带入，h_1, h_2, h_3均$= 0$，则$o = f(-1) = 0$。

当任意class为1的样本带入，h_1, h_2, h_3至少有一个值$= 1$(其他为0)，则$o = f(x \geq 0) = 1$。

因此通过取这样的参数，我们的神经网络可以100%的将数据成功划分。

$$\begin{bmatrix} w_{0,1}^{[1]} & w_{1,1}^{[1]} & w_{2,1}^{[1]} \\ w_{0,2}^{[1]} & w_{1,2}^{[1]} & w_{2,2}^{[1]} \\ w_{0,3}^{[1]} & w_{1,3}^{[1]} & w_{2,3}^{[1]} \end{bmatrix} = \begin{bmatrix} 0.4 & -1 & 0 \\ 0.3 & 0 & -1 \\ -4.5 & 1 & 1 \end{bmatrix}$$
$$\begin{bmatrix} w_0^{[2]} & w_1^{[2]} & w_2^{[2]} & w_3^{[2]} \end{bmatrix} = [-1 \quad 1 \quad 1 \quad 1]$$

(c)

$\because h_1, h_2, h_3$的激活函数是线性函数$f(x) = x$，而仅仅output o的激活函数是step function。所以输入样本$x^{(i)}$ 通过整个神经网络输出$o^{(i)}$的式子可以表示为：

$o^{(i)} = f((w^{[2]})^T((w^{[1]})^T x^{(i)}))$，其中$w^{[1]}$ 为从$x^{(i)}$到hidden neuron的权重矩阵，$w^{[2]}$ 为从hidden neuron到输出$o^{(i)}$的权重矩阵，f是step function。

所以可以看出，step function内部是一个与$x^{(i)}$ 相关的线性函数，再通过step function之后，只能够将内部计算出来≥ 0的变成1，将内部计算出来< 0的变成0。

而对于题中的样本，通过一条直线是无法将两个类别完全划分开来的，因此$o^{(i)} = f((w^{[2]})^T((w^{[1]})^T x^{(i)}))$无法将两个类别100%识别正确，那么loss也不会为0。

因此，对于"h_1, h_2, h_3的激活函数是线性函数$f(x) = x$，而仅仅output o的激活函数是step function"这种情况，是无法找到一组权重能够让loss为0的。