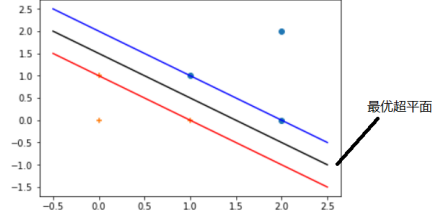


1 Homework4.2 Neural Network

解： (a) 依题意可得： 令 w_1 标签为 -1 , w_2 标签为 $+1$, 即 $z_1 = z_2 = z_3 = -1$, $z_4 = z_5 = z_6 = +1$ 由图可得最优超平面为 $(3/2 - 1 - 1)(1y_1y_2) = 0$ 为了使 $z_k \mathbf{a} \mathbf{y} \geq 1$, 权重向



量该为 $(3 - 2 - 2)^t$, 而最优间隔为 $\sqrt{2}/4$

(b) 支持向量为距离超平面最近的点, 由此可知本题所述的支持向量为: $x_1 = (1, 1)^t$; $x_3 = (2, 0)^t$; $x_5 = (1, 0)^t$; $x_6 = (0, 1)^t$

(c) 最大化拉格朗日函数 $L(\alpha)$ 其中

$$L(\alpha) = \sum_{k=1}^n \alpha_k - \frac{1}{2} \sum_{k,j} \alpha_k \alpha_j z_k z_j \mathbf{y}_j^t \mathbf{y}_k \quad (1)$$

约束条件为:

$$\sum_{k=1}^n z_k \alpha_k = 0 \quad (2)$$

其中 $\alpha_k \geq 0$ 基于上述约束条件, 有 $\alpha_6 = \alpha_1 + \alpha_2 + \alpha_3 - \alpha_4 + \alpha_5$. 令

$$\frac{\partial L(0, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6)}{\partial \alpha_i} = 0 \quad (3)$$

此时 $\alpha = 1/5(0, -2, -2, 8, -8, -4)^t$ 不满足条件。而

$$\frac{\partial L(\alpha_1, 0, \alpha_3, \alpha_4, \alpha_5)}{\partial \alpha_i} = \frac{\partial L(\alpha_1, \alpha_2, 0, \alpha_4, \alpha_5)}{\partial \alpha_i} = 0 \quad (4)$$

方程不相容。再令

$$\frac{\partial L(\alpha_1, \alpha_2, \alpha_3, 0, \alpha_5)}{\partial \alpha_i} = 0 \quad (5)$$

得到 $\alpha = 1/5(16, 0, 4, 0, 14, 6)^t$ 满足条件, 此时 $L(\alpha) = 4$ 最后令

$$\frac{\partial L(\alpha_1, \alpha_2, \alpha_3, \alpha_4, 0)}{\partial \alpha_i} = 0 \quad (6)$$

得到 $\alpha = 1/5(2, 2, 2, 0, 0, 6)^t$ 满足条件, 此时 $L(\alpha) = 1.2$ 由此可得 $\alpha = 1/5(16, 0, 4, 0, 14, 6)^t$ 。

现在求权重向量 a 。即最大化对偶空间目标函数，对偶空间目标函数对 a 求导可得：

$$\frac{\partial L}{\partial a} = a - \sum_{k=1}^n \alpha_k z_k y_k = 0 \quad (7)$$

将 $\alpha = 1/5(16, 0, 4, 0, 14, 6)^t$ 代入上式可得 $a = (0, -2, -2)$

由于 $\|a\|^2$ 并没有包含向量 a ，而且 $\sum_k a_k z_k = 0$ 。为了求 a_0 通过使用其中一个支持向量 $y_1 = (1, 1, 1)^t$ ，由于有 $a^t y_1 z_1 = 1$ ，因此 $-(0, -2, -2)^t (1, 1, 1) = -a_0 + 4 = 1$ 。由此可得 $a_0 = 3$

由此可得全部的权重向量为 $a = (3, -2, -2)^t$

2 Homework4.2 Neural Network

解：（a）依题意可得：隐含层 h_1, h_2, h_3 分别为：

$$h_1 = w_{1,1}^{[1]} x_1 + w_{2,1}^{[1]} x_2 + w_{0,1}^{[1]}; \quad (8)$$

$$h_2 = w_{1,2}^{[1]} x_1 + w_{2,2}^{[1]} x_2 + w_{0,2}^{[1]}; \quad (9)$$

$$h_3 = w_{1,3}^{[1]} x_1 + w_{2,3}^{[1]} x_2 + w_{0,3}^{[1]}; \quad (10)$$

sigmoid 激活函数的数学表达式为：

$$s(t) = \frac{1}{1 + e^{-t}} \quad (11)$$

于是可得输出为：

$$o = s(w_1^{[2]} s(h_1) + w_2^{[2]} s(h_2) + w_3^{[2]} s(h_3) + w_0^{[2]}) \quad (12)$$

损失函数为：

$$l = \frac{1}{m} \sum_{i=1}^m (o^{(i)} - y^{(i)})^2 \quad (13)$$

由题意可知学习率为 α 利用梯度下降法求参数 $w_{1,2}^{[1]}$ 的更新方式为：

$$w_{1,2}^{[1]} = w_{1,2}^{[1]} - \alpha \frac{\partial l}{\partial w_{1,2}^{[1]}} \quad (14)$$

将 h_1, h_2, h_3 代入 o 中，再将 o 代入 l 中求导可得：

$$\frac{\partial l}{\partial w_{1,2}^{[1]}} = \frac{2}{m} \sum_{i=1}^m (o^{(i)} - y^{(i)}) \frac{\partial o}{\partial w_{1,2}^{[1]}} \quad (15)$$

利用 python 计算 $\frac{\partial o}{\partial w_{1,2}^{[1]}}$

```
In [2]: from sympy import *

# sigmoid 函数
def sh(h):
    out = 1/(1+exp(-h))
    return out

# 设置参数
w11, x1, w21, x2, w01, w12, w22, w02, w13, w23, w03 = symbols('w11 x1 w21 x2 w01 w12 w22 w02 w13 w23 w03')
# 隐含层
h1 = w11 * x1 + w21 * x2 + w01
h2 = w12 * x1 + w22 * x2 + w02
h3 = w13 * x1 + w23 * x2 + w03

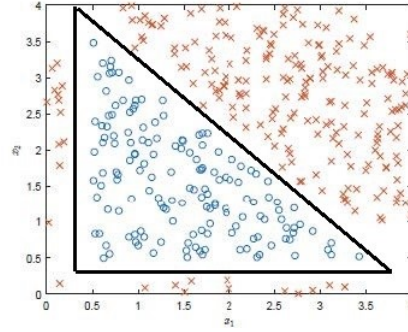
w1, w2, w3, w0 = symbols('w1 w2 w3 w0')
o = w1/sh(h1) + w2/sh(h2) + w3/sh(h3) + w0
so = sh(o)
# 打印输出
diff_so = diff(so, w12)
print(diff_so)

-w2*x1*exp(-w02 - w12*x1 - w22*x2)*exp(-w0 - w1*(exp(-w01 - w11*x1 - w21*x2) + 1) - w2*(exp(-w02 - w12*x1 - w22*x2) + 1) - w3*(exp(-w03 - w13*x1 - w23*x2) + 1))/(exp(-w0 - w1*(exp(-w01 - w11*x1 - w21*x2) + 1) - w2*(exp(-w02 - w12*x1 - w22*x2) + 1) - w3*(exp(-w03 - w13*x1 - w23*x2) + 1)) + 1)**2
```

于是可得 $w_{1,2}^{[1]}$ 的更新方式为: (其中 diff_so 由上式给出.)

$$w_{1,2}^{[1]} = w_{1,2}^{[1]} - \alpha \text{diff_so} \quad (16)$$

(b) 在激活函数为 step function 的情况下, 可以用三条线性函数对数据分类, 划分方式如图一



其中三条线分别为: $x_1 = 0.25; x_2 = 0.25; x_1 + x_2 = 4.3$, 也就是说数据满足条件 $x_1 \leq 0.25, x_2 \leq 0.25, x_1 + x_2 \geq 4.3$ 时, 此时分类为 1, 其余分类为 0. 由此可推出各个权重参数的一个组合为:

$$w_{0,1}^{[1]} = 1, w_{1,1}^{[1]} = 1, w_{2,1}^{[1]} = -4.3 \quad (17)$$

$$w_{0,2}^{[1]} = 0, w_{1,2}^{[1]} = 4, w_{2,2}^{[1]} = -1.3 \quad (18)$$

$$w_{0,3}^{[1]} = 4, w_{1,3}^{[1]} = 0, w_{2,3}^{[1]} = -1.3 \quad (19)$$

$$w_0^{[2]} = 1, w_1^{[2]} = 1, w_2^{[2]} = 1, w_3^{[2]} = 0 \quad (20)$$

(c) 如果激活函数为一个线性函数 $f(x) = x$, 不存在一个权重组合使得损失函数为 0. 在激活函数为线性函数的情况下, 题中所述的神经网络其实还是一个线性函数, 而线性函数是不能够将题中所述的两类数据完全分开的。如下图所示。

