姓名：李千鹏 学号：2021E8014682047 院系：人工智能学院

问题 1.

对样本进行规范化增广，得到样本矩阵为：

$$Y = \begin{pmatrix} 1 & 2 & -4 & -3 \\ 4 & 3 & -1 & -2 \\ 1 & 1 & -1 & -1 \end{pmatrix}$$

第一次迭代, $a_0^T Y = (4 \quad 3 \quad -1 \quad -2)$，可知第三个第四个样本没有正确分类

则 $a_1 = a_0 + Y_{*3} + Y_{*4} = (-7 \quad -2 \quad -2)^\top$

第二次迭代，$a_1^\top Y = (-17 \quad -22 \quad 32 \quad 27)$，可知第一个第二个样本没有正确分类

则 $a_2 = a_1 + Y_{*1} + Y_{*2} = (-4 \quad 5 \quad 0)^\top$

第三次迭代，$a_2^\top Y = (16 \quad 7 \quad 11 \quad 2)$，可知所有样本全部正确分类

所以最终 $a = (-4 \quad 5 \quad 0)^\top$

问题 2.

当预测类别为 $w1$ 时，$\begin{cases} g_1(x) > g_2(x) \\ g_1(x) > g_3(x) \end{cases} \Rightarrow \begin{cases} x_2 > \frac{x_1}{2} \\ x_1 < \frac{1}{2} \end{cases}$

决策面为：$\begin{cases} 2x_2 - x_1 = 0 \\ x_1 = \frac{1}{2} \end{cases}$

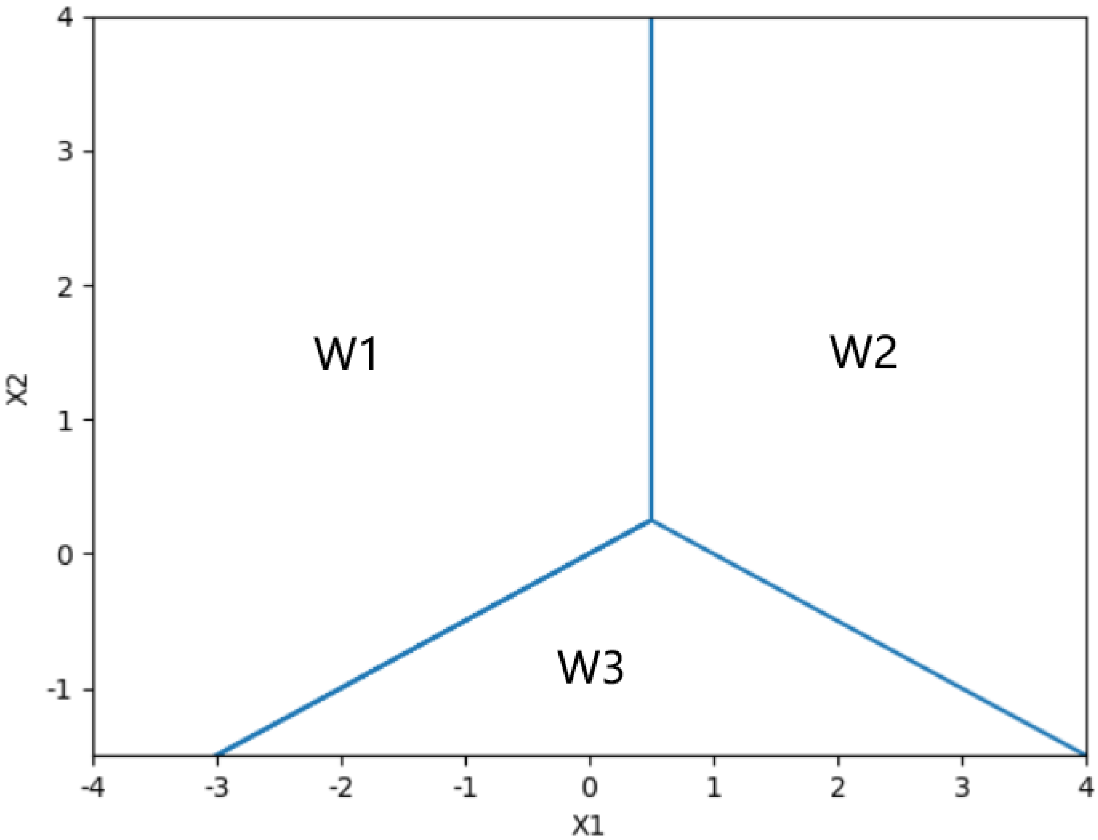
当预测类别为 $w2$ 时，$\begin{cases} g_2(x) > g_1(x) \\ g_2(x) > g_3(x) \end{cases} \Rightarrow \begin{cases} x_1 > \frac{1}{2} \\ x_2 > \frac{1}{2} - \frac{x_1}{2} \end{cases}$

决策面为：$\begin{cases} x_1 = \frac{1}{2} \\ 2x_2 + x_1 - 1 = 0 \end{cases}$

当预测类别为 $w3$ 时，$\begin{cases} g_3(x) > g_1(x) \\ g_3(x) > g_2(x) \end{cases} \Rightarrow \begin{cases} x_2 < \frac{x_1}{2} \\ x_2 < \frac{1}{2} - \frac{x_1}{2} \end{cases}$

决策面为：$\begin{cases} x_2 = \frac{x_1}{2} \\ 2x_2 + x_1 - 1 = 0 \end{cases}$

绘制出所有的决策面，如下图



通过该图可以很明显的看出，各类决策区域的并集为整个平面，即不存在分类不确定区域。

问题 3.

(1)

编写代码求解 *batch perception* 算法下，$w1-w2$、$w3-w2$ 分类的收敛步数。算法参数 $\theta=0.00001, \eta=0.001$。$\eta$ 为学习率，θ 为 $\eta*$ 梯度的阈值。

结果如图所示

w1与w2分类迭代次数: 24
w3与w2分类迭代次数: 17

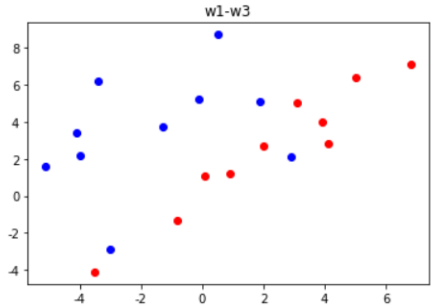
(2)

编写 *Ho − Kashyap algorithm*，获得 $w1-w3$、$w2-w4$ 训练迭代次数、准确度以及误差 e。

算法最大迭代次数 $iter\_max = 10000$，学习率 $\eta = 0.1$，$w1 - w3$ 训练结果如下。

```
acc: 0.85
e:
[[-3.83417001e-01]
 [-1.49184173e-01]
 [-5.65745479e-01]
 [-1.78122712e-02]
 [-1.92845053e-01]
 [-7.18926724e-02]
 [-9.33265232e-03]
 [-1.57615956e-01]
 [-5.68106756e-02]
 [ 0.00000000e+00]
 [-6.63432043e-01]
 [ 4.44089210e-16]
 [-7.42253796e-01]
 [ 2.22044605e-16]
 [ 4.44089210e-16]
 [ 2.22044605e-16]
 [ 4.44089210e-16]
 [ 4.44089210e-16]
 [ 4.44089210e-16]
 [-1.98970094e-01]]
w1与w3分类迭代次数:  10000
```

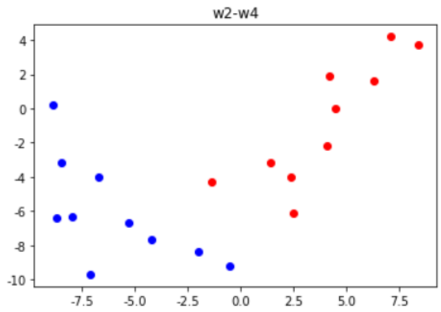
在 $w1 - w3$ 分类中，算法达到最大迭代次数仍不能够正确分类。做出两类数据的散点图，发现其线性不可分，故算法不会收敛。



$w2 - w4$ 训练结果如下

```
acc: 1.0
e:
[[ 0.01228558]
 [-0.32853785]
 [ 0.00828874]
 [ 0.01031623]
 [ 0.00917616]
 [ 0.00454559]
 [ 0.00471684]
 [ 0.00359158]
 [ 0.01281359]
 [ 0.00679508]
 [ 0.00051573]
 [-0.09596537]
 [ 0.00149195]
 [ 0.00162886]
 [ 0.00095866]
 [-0.17094952]
 [-0.004754  ]
 [ 0.00356332]
 [ 0.00443171]
 [ 0.00307019]]
w2与w4分类迭代次数:  599
```

对 $w2 - w4$ 这两类数据进行训练，算法迭代了 *599* 次就实现了收敛，做出 $w2 - w4$ 两类数据的散点图，发现其线性可分，故在有限步下能够实现收敛。



(3)

MSE 多类扩展主要涉及的是权重矩阵 W 的计算：$\hat{\mathbf{W}} = \left(\hat{\mathbf{X}}\hat{\mathbf{X}}^T + \lambda\mathbf{I}\right)^{-1}\hat{\mathbf{X}}\mathbf{Y}^T$

式中 $\hat{\mathbf{X}}$ 为增广样本矩阵，$\hat{\mathbf{X}} = (\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \cdots, \hat{\mathbf{x}}_n) \in R^{(d+1)\times n}$

$\hat{\mathbf{Y}}$ 为构造的回归值，$\hat{\mathbf{Y}} \in R^{(c)\times n}$

λ 为一个很小的正数，本例 $\lambda = 0.0000005$

分类样本时，若 $j = \arg\max\left(\mathbf{W}^T\hat{\mathbf{x}}\right)$，则 $\hat{\mathbf{x}} \in \omega_j$

将每一类的前 *8* 个样本用于训练，后 *2* 个样本用于测试，分类结果如下

$acc = 1.0$