一、　线性判别函数的学习可以采用多种不同的学习算法。（**20** 分）

a) **请分别说明感知器算法和最小均方误差算法所采用的优化准则函数，并说明两种算法的不同特点。（8 分）**

b) **现有两类训练样本：（8 分）**

$$\omega_1\text{：}\left\{(1,1)^t,(0,1)^t\right\}\text{，}\omega_2\text{：}\left\{(-1,0)^t,(0,-1)^t\right\}$$

请用单样本调整的感知器算法学习线性分类器，初始权矢量为 $\mathbf{a}(0)=(1,2,-2)^t$，

a 的第 1 维为偏置，学习率 $\eta=1$

c) **画出上述线性分类器的判别界面。（4 分）**

答：a) 感知器算法以错分样本到判别界面距离之和作为准则:

$$J_P(\mathbf{w})=\sum_{\mathbf{y}\in Y}\left(-\mathbf{a}^t\mathbf{y}\right)$$

最小均方误差算法将不等式组的求解转化为方程组的求解，以解得误差矢量的长度平方最小为准则：

$$J_S(\mathbf{w})=\left\|\mathbf{Ya}-\mathbf{b}\right\|^2$$

感知器算法的特征是当样本集合线性可分时算法收敛，但样本不可分时算法无法收敛，一般来说算法收敛速度较慢。

最小均方误差算法当样本不可分时也能收敛于均方误差最小解，当样本数区域无穷时以均方误差逼近贝叶斯判别函数。

b) 样本规范化：$\mathbf{y}_1=(1,1,1)^t$，$\mathbf{y}_2=(1,0,1)^t$，$\mathbf{y}_3=(-1,1,0)^t$，$\mathbf{y}_4=(-1,0,1)^t$

第 1 轮：

$\mathbf{a}^t(0)\mathbf{y}_1=1>0$，不需修改；

$\mathbf{a}^t(0)\mathbf{y}_2=-1<0$，$\mathbf{a}(1)=\mathbf{a}(0)+\mathbf{y}_2=(2,2,-1)^t$

$\mathbf{a}^t(1)\mathbf{y}_3=0$，$\mathbf{a}(2)=\mathbf{a}(1)+\mathbf{y}_3=(1,3,-1)^t$

$\mathbf{a}^t(2)\mathbf{y}_4=-2<0$，$\mathbf{a}(3)=\mathbf{a}(2)+\mathbf{y}_4=(0,3,0)^t$

第 2 轮：

$\mathbf{a}^t(3)\mathbf{y}_1=3>0$

$\mathbf{a}^t(3)\mathbf{y}_2=0$，$\mathbf{a}(4)=\mathbf{a}(3)+\mathbf{y}_2=(1,3,1)^t$

$\mathbf{a}^t(4)\mathbf{y}_3=2>0$

$\mathbf{a}^t(4)\mathbf{y}_4=0$，$\mathbf{a}(5)=\mathbf{a}(4)+\mathbf{y}_4=(0,3,2)^t$

第 3 轮：

$\mathbf{a}^t(5)\mathbf{y}_1=5>0$，$\mathbf{a}^t(5)\mathbf{y}_2=2>0$，$\mathbf{a}^t(5)\mathbf{y}_3=3>0$，$\mathbf{a}^t(5)\mathbf{y}_4=2>0$

输出：$\mathbf{a}=(0,3,2)^t$

d) 判别函数为：$g(\mathbf{x})=3x_1+2x_2$