第一题

将各样本的齐次坐标进行规范化处理后，得 $\mathbf{y}_1 = (1,\ 4,\ 1)^T$, $\mathbf{y}_2 = (2,\ 3,\ 1)^T$, $\mathbf{y}_3 = (-4,\ -1,\ -1)^T$, $\mathbf{y}_4 = (-3,\ -2,\ -1)^T$.
准则函数为，

$$J(\mathbf{a}) = \sum_{\mathbf{y}\in Y}(-\mathbf{a}^T\mathbf{y})$$

其中 Y 为错分样本的集合，优化目标是最小化 $J(\mathbf{a})$。将 $J(\mathbf{a})$ 对 $\mathbf{a}$ 求导，得，

$$\frac{\partial J(\mathbf{a})}{\partial \mathbf{a}} = \sum_{\mathbf{y}\in Y} -\mathbf{y}$$

梯度下降法为每一步向负梯度方向移动一段距离，即，

$$\mathbf{a}_{k+1} = \mathbf{a}_k + \eta_k \sum_{\mathbf{y}\in Y}\mathbf{y}$$

当 $k=1$ 时，$\mathbf{a}_1^T\mathbf{y}_1 = 4$, $\mathbf{a}_1^T\mathbf{y}_2 = 3$, $\mathbf{a}_1^T\mathbf{y}_3 = -1$, $\mathbf{a}_1^T\mathbf{y}_4 = -2$, . 其中错分样本为 $\mathbf{y}_3$、$\mathbf{y}_4$，代入公式得，

$$\mathbf{a}_2 = \mathbf{a}_1 + \mathbf{y}_3 + \mathbf{y}_4 = (-7,\ -2,\ -2)^T$$

当 $k=2$ 时，$\mathbf{a}_2^T\mathbf{y}_1 = -17$, $\mathbf{a}_2^T\mathbf{y}_2 = -22$, $\mathbf{a}_2^T\mathbf{y}_3 = 32$, $\mathbf{a}_2^T\mathbf{y}_4 = 27$, . 其中错分样本为 $\mathbf{y}_1$、$\mathbf{y}_2$，代入公式得，

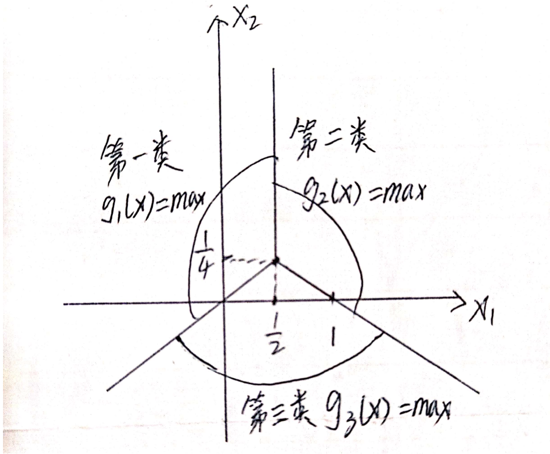
$$\mathbf{a}_3 = \mathbf{a}_1 + \mathbf{y}_1 + \mathbf{y}_2 = (-4,\ 5,\ 0)^T$$

当 $k=3$ 时，$\mathbf{a}_3^T\mathbf{y}_1 = 16$, $\mathbf{a}_3^T\mathbf{y}_2 = 7$, $\mathbf{a}_3^T\mathbf{y}_3 = 11$, $\mathbf{a}_3^T\mathbf{y}_4 = 2$, . 没有错分样本，跳出循环，返回最终结果为，

$$\mathbf{a} = (-4,\ 5,\ 0)^T$$

第二题

决策面图像为:



由上图可知，3 个决策区域不相交，所以不存在分类不确定的区域。

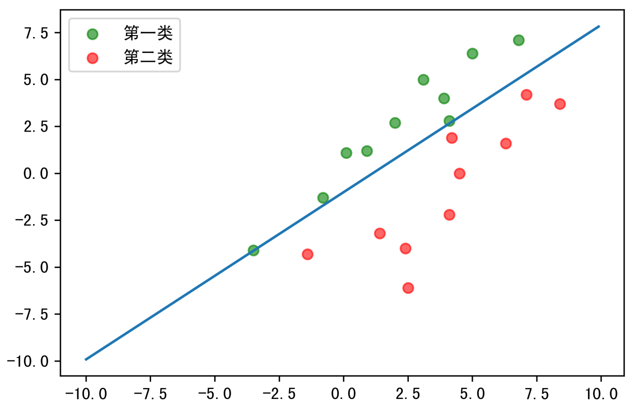
第三题

(1) 使用批处理感知器算法求解线性判别函数权向量 $\mathbf{a}$，步长因子 η 设置为 1，初始化权重向量为 $\mathbf{a}=\mathbf{0}$。

(a) 类别 ω_1 与 ω_2 经过 23 次迭代后得到结果。

$$\mathbf{a}=[-30.4,\ 34.1,\ 34.0]^T$$

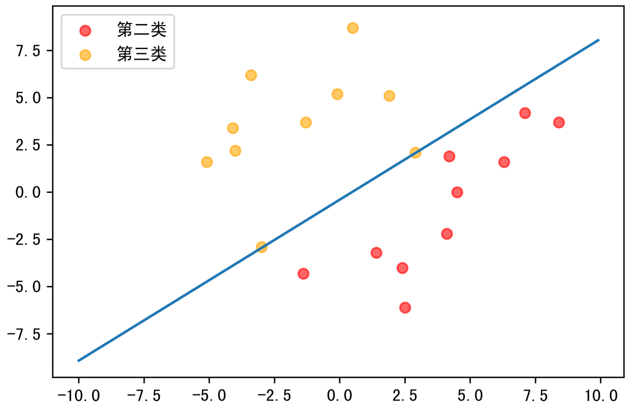
可视化如下，



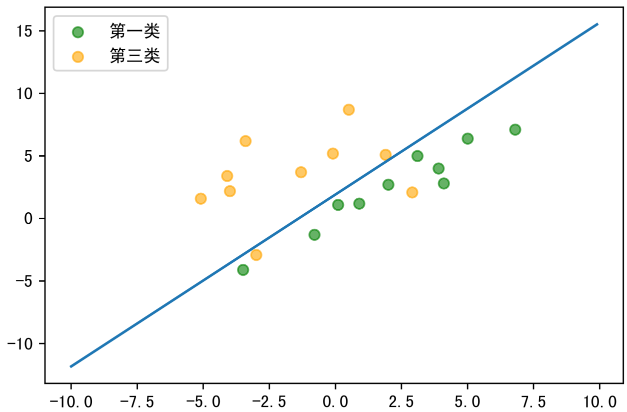
(b) 类别 ω_2 与 ω_3 经过 16 次迭代后得到结果。

$$\mathbf{a} = [41.4,\ -48.6,\ -19.0]^T$$

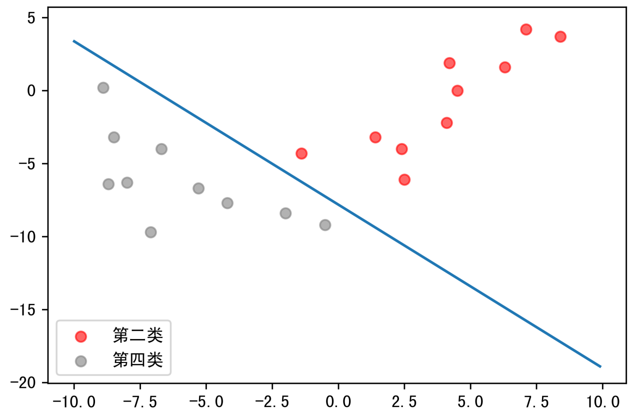
可视化如下，



(2) 设置步长为 1，$\mathbf{a} = \mathbf{0}$，$\mathbf{b} = \mathbf{1}$。$b_{min} = 0.0001$，$k_{max} = 10000$。当使用样本 ω_1 和 ω_3 时，无法在 k_{max} 限制内收敛，有两个错分样本，错误率为 10%，误差为 8.451。可视化如下，且由图可知，样本线性不可分。



当使用样本 ω_2 和 ω_4 时，迭代 667 次后收敛，无错分样本，错误率为 0%，误差为 1.384×10^{-8}。可视化如下，



(3) 根据 MSE 多类扩展方法，得

$$\hat{\mathbf{W}} = (\hat{\mathbf{X}}\hat{\mathbf{X}}^T + \lambda\mathbf{I})^{-1}\hat{\mathbf{X}}\hat{\mathbf{Y}}^T$$

其中，$\hat{\mathbf{X}}$ 的每一列是样本点的增广坐标，加 $\lambda\mathbf{I}$ 是为了使该项可逆，设置

$\lambda = 0.0001$。

在训练样本上的准确率为 71.875%，在测试样本上的准确率为 100%。

具体代码及数据详见 temp.py