



目录

CONTENTS

- 1 实验设置
- 2 Iris数据集画图看选择样本的不同
- 3 观察加入SPL与不加入选择样本的不同
- 4 真实数据集的准确率变化结果



实验设置

模型：Linear Regression $y=X*w$ (X的最后加了一列1向量)

$$\min_{Q,f} \sum_{\{x_i, y_i\} \in L} (y_i - f(x_i))^2 + \sum_{x_i \in Q} (\hat{y}_i - f(x_i))^2 + \mu MMD(S, L \cup Q) + \gamma \|f\|^2$$

$$\min_{Q,f} \sum_{\{x_i, y_i\} \in L} \alpha_i (y_i - f(x_i))^2 + \lambda \left(\frac{1}{2} \alpha^2 - \alpha \right) + \sum_{\{x_i, y_i\} \in Q} \beta_i (\hat{y}_i - f(x_i))^2 + \lambda \left(\frac{1}{2} \beta^2 - \beta \right) + \mu MMD(S, L \cup Q) + \gamma \|f\|^2$$

$$\min_{f, \theta} \sum_{\{x_i, y_i\} \in L \cup Q} \alpha_i (y_i - f(x_i))^2 + \lambda \left(\frac{1}{2} \alpha^2 - \alpha \right) + \sum_{x_j \in U} \theta_j * \beta_j (\hat{y}_j - f(x_j))^2 + \lambda \left(\frac{1}{2} \beta^2 - \beta \right) + \mu MMD(S, L \cup Q) + \gamma \|f\|^2$$

Repeat:

Repeat:

1. Solve SPL for the unlabeled data to get θ and β $\arg \min_{\beta} \sum_{x_j \in U} \theta_j * \beta_j (\hat{y}_j - f(x_j))^2 + \lambda \left(\frac{1}{2} \beta^2 - \beta \right)$
2. Solve QP for the unlabeled data to get θ $\arg \min_{\theta} \sum_{x_j \in U} \theta_j * \beta_j (\hat{y}_j - f(x_j))^2 + \mu MMD(S, L \cup Q)$

Until Converge

3. Pick the largest θ b instances to Query

4. Calculate the weight of set L and Q using the true label. $\sum_{\{x_i, y_i\} \in L \cup Q} \alpha_i (y_i - f(x_i))^2 + \lambda \left(\frac{1}{2} \alpha^2 - \alpha \right)$

5. Optimize the weighted loss (cvx) $\sum_{\{x_i, y_i\} \in L \cup Q} \alpha_i (y_i - f(x_i))^2 + \gamma \|f\|^2$

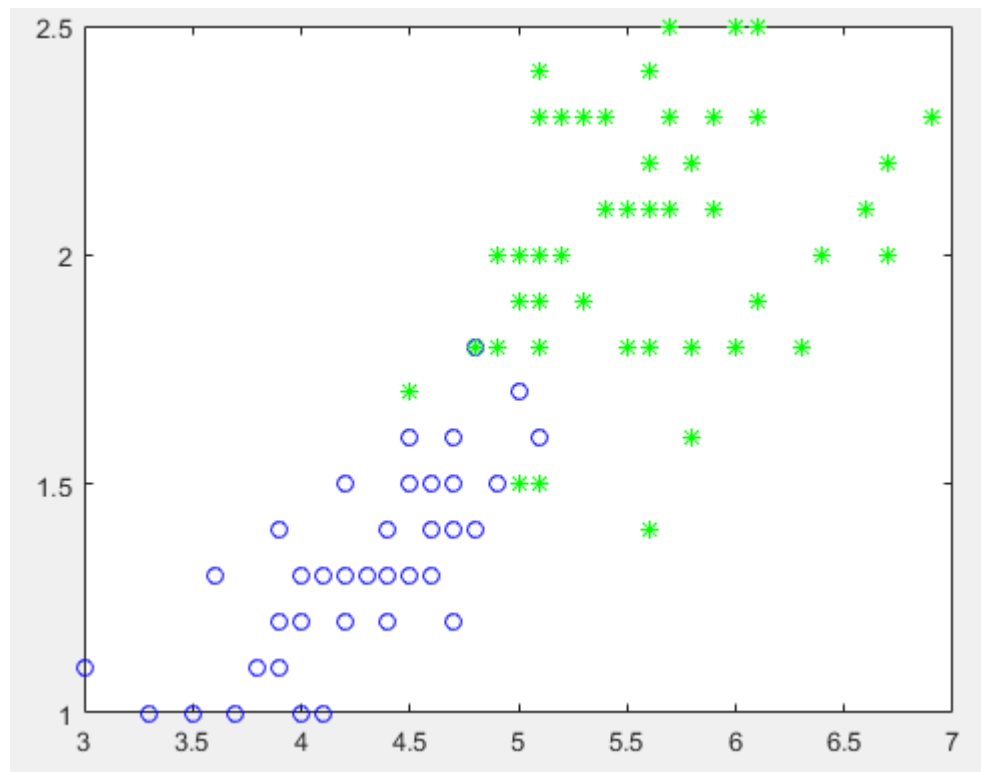
6. increase λ

Until 80% samples are labeled

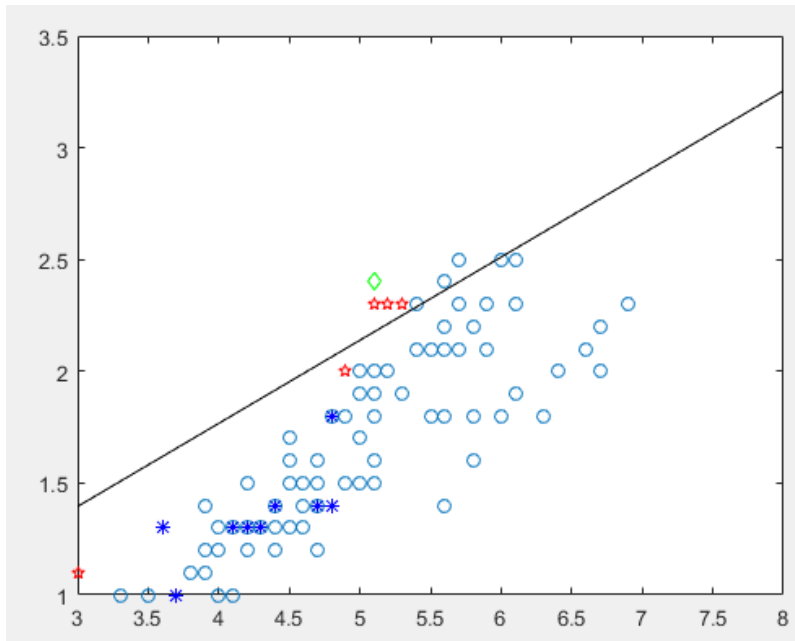
$$\lambda_j^t = \begin{cases} \lambda_j^0, & t = 0 \\ \lambda_j^{(t-1)} + \alpha * \eta_j^t, & 1 \leq t \leq \tau \\ \lambda_j^{(t-1)}, & t > \tau, \end{cases}$$



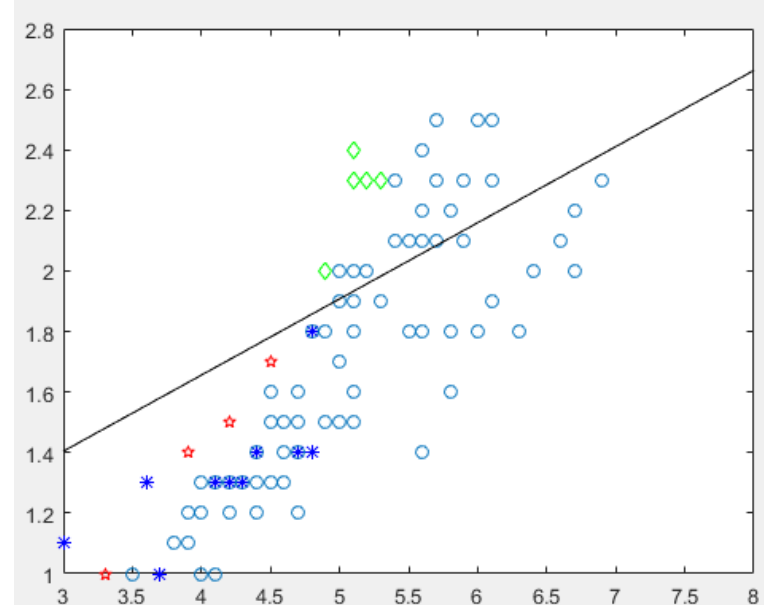
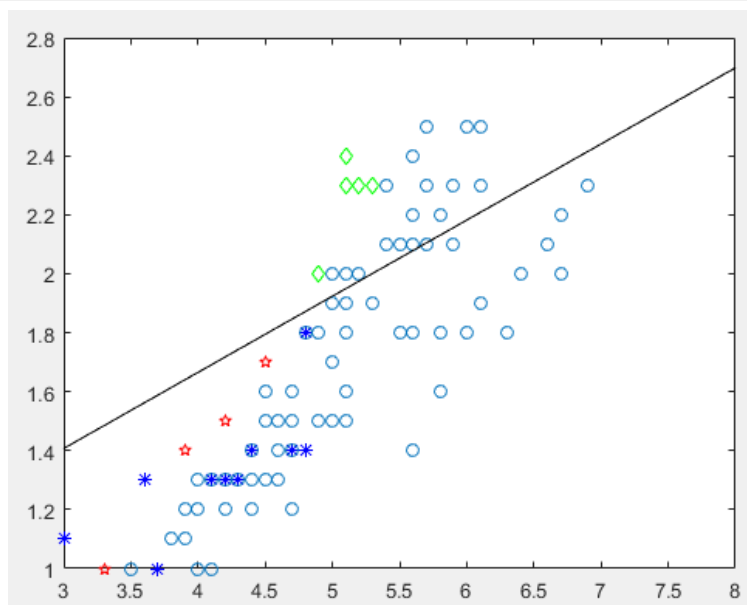
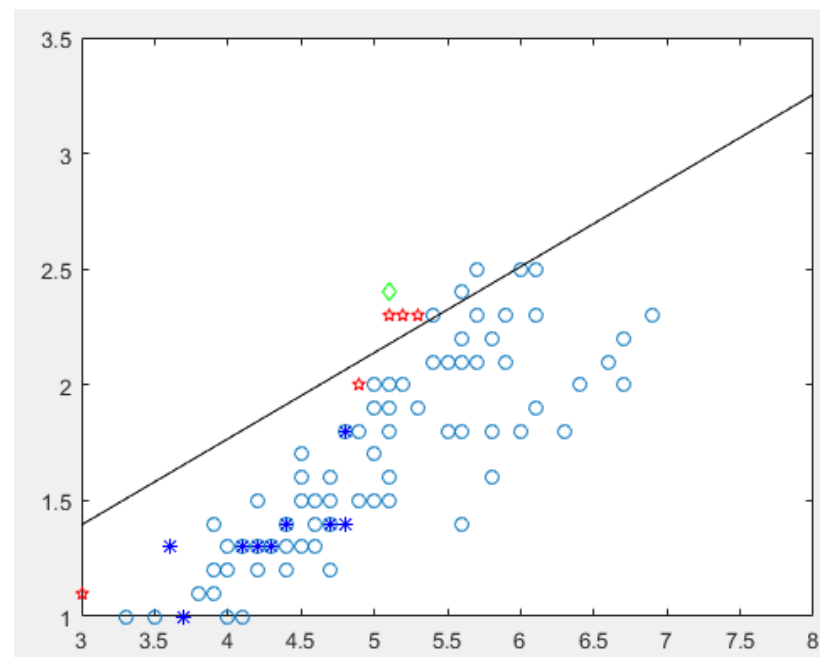
Iris数据集画图看选择样本的不同



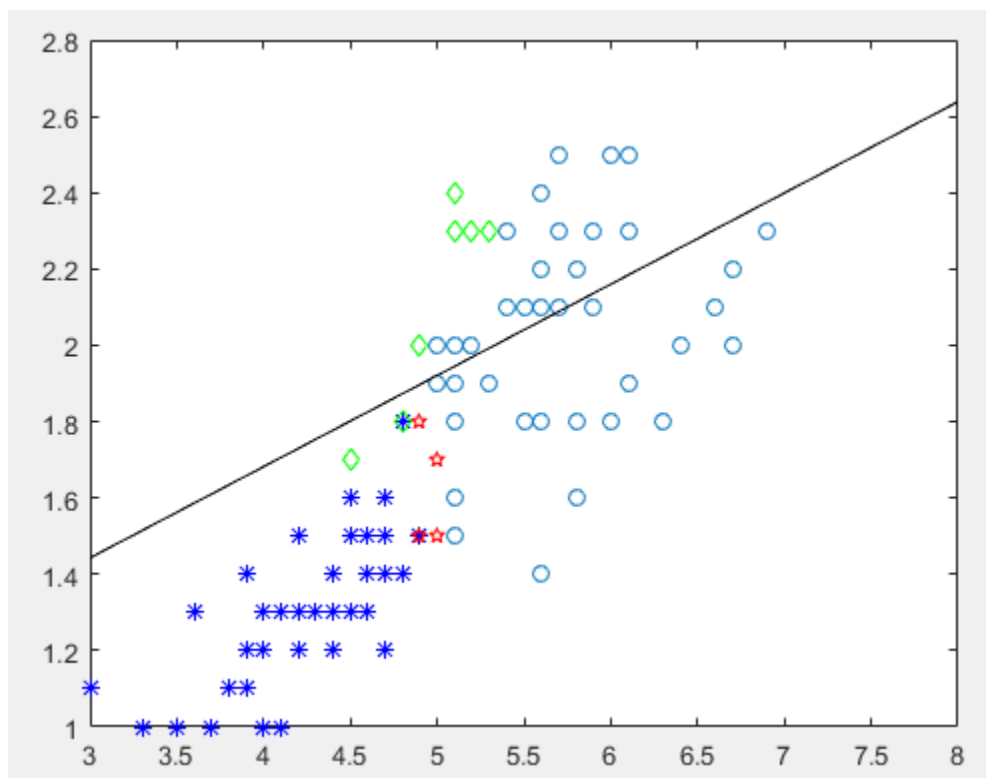
SPL $\lambda = 0.5, \text{pace} = 0.1$



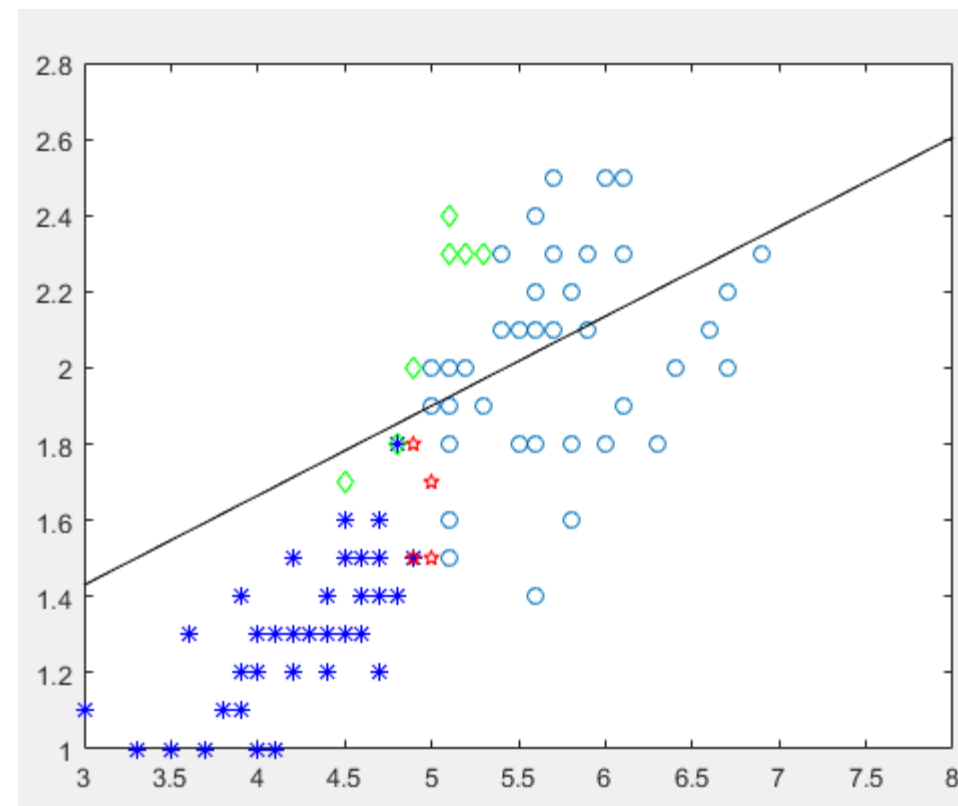
noSPL

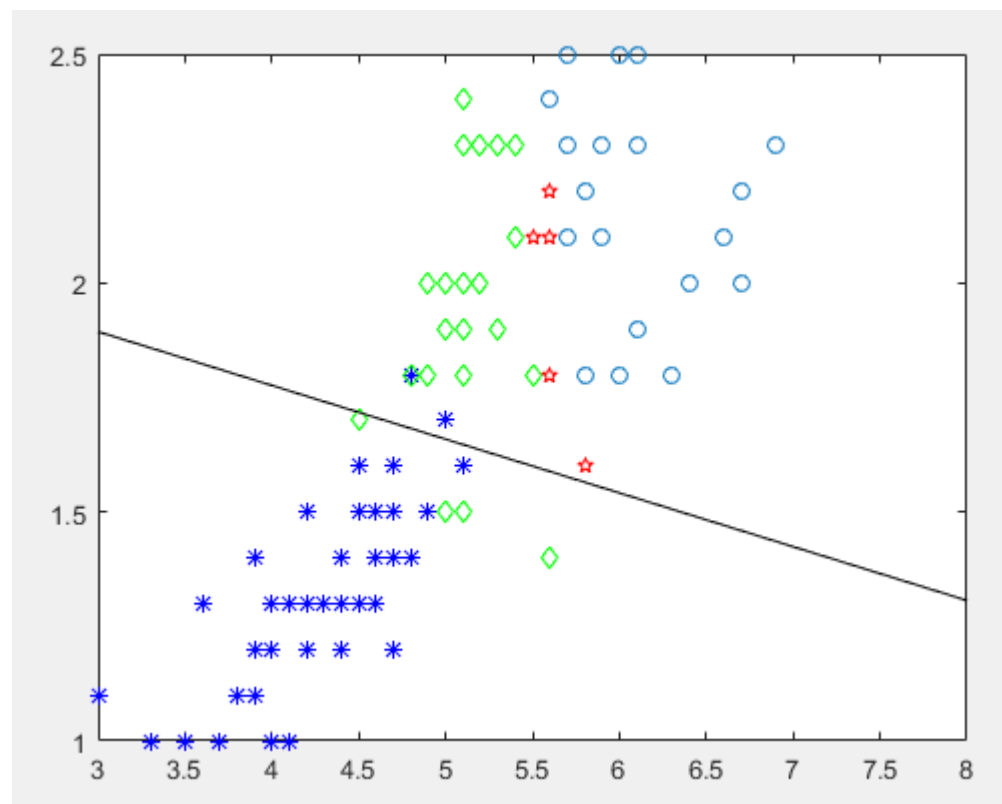


SPL

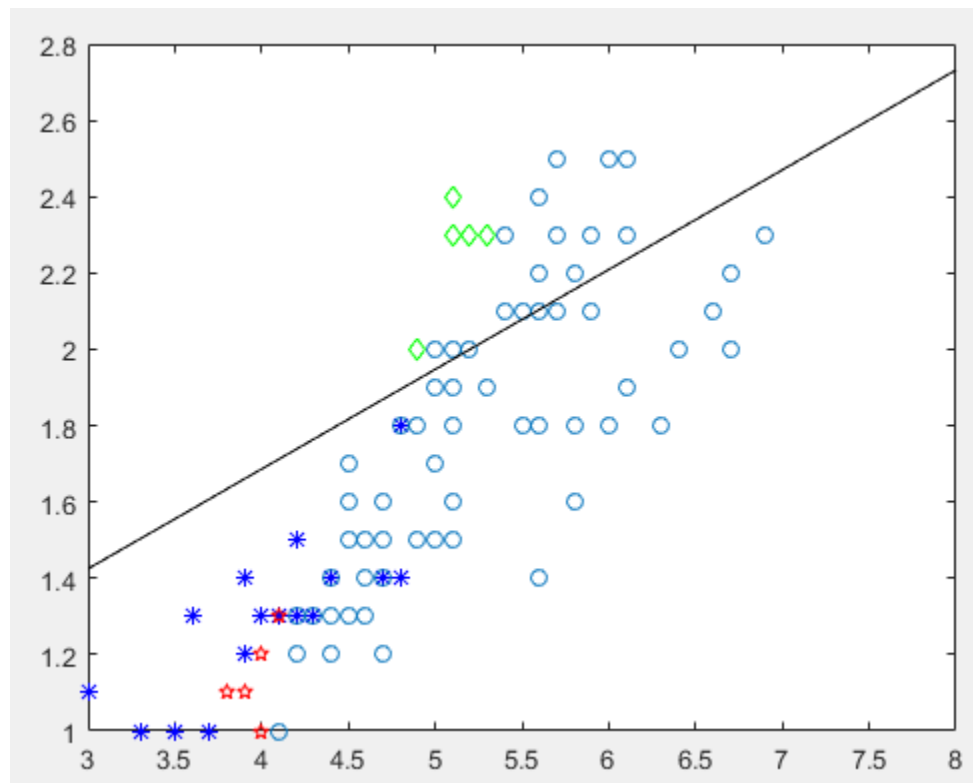


noSPL



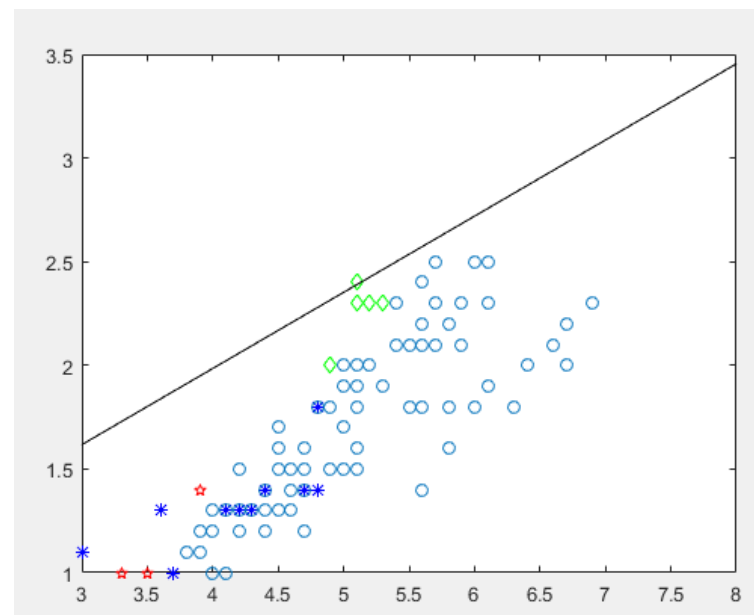
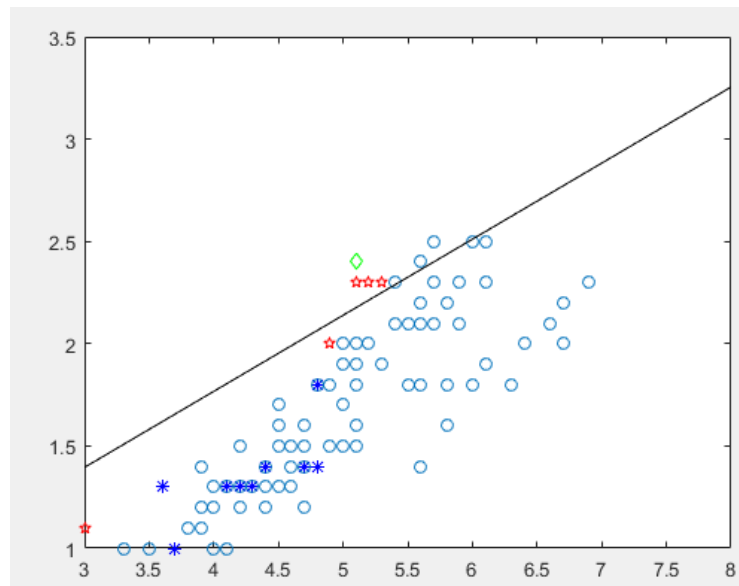


SPL $\lambda = 0.2, \text{pace} = 0.1$

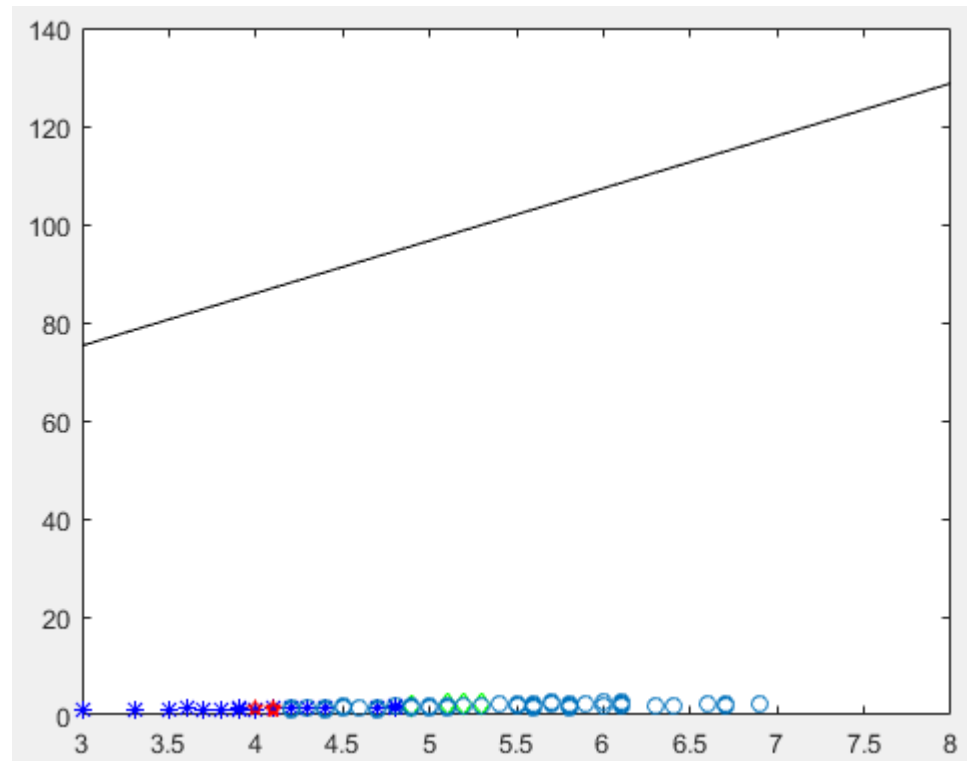
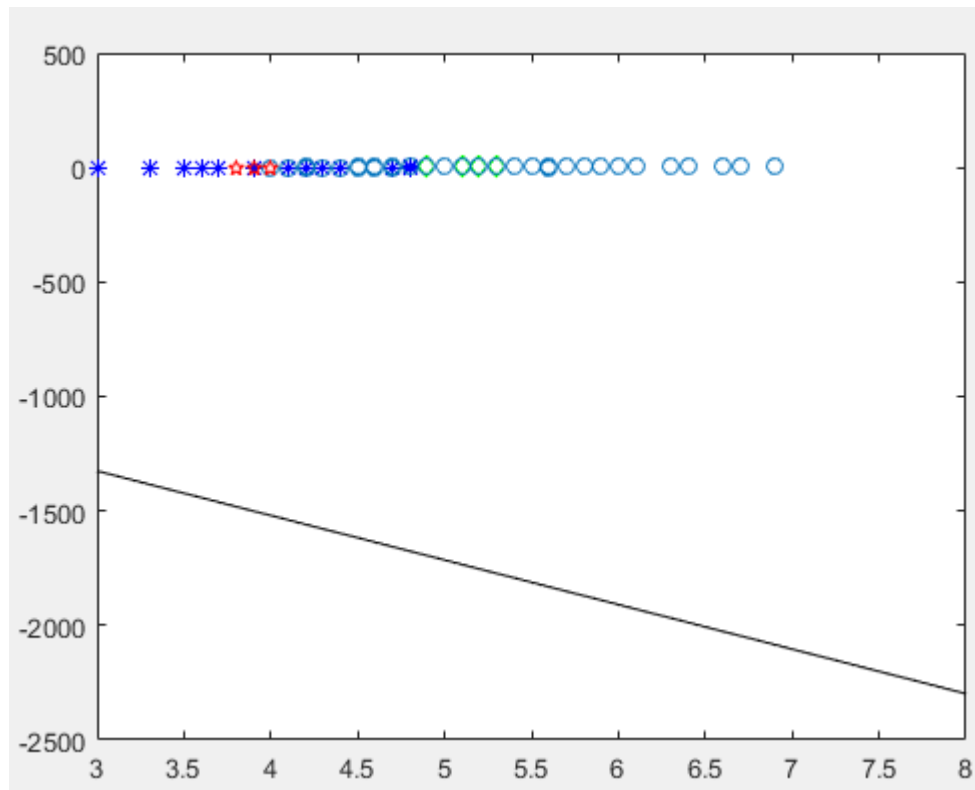


Iter = 6 , 与之前的差不多

SPL $\lambda = 0.15, \text{pace} = 0.08$



SPL $\lambda = 0.15, \text{pace} = 0.08$



交替优化 $\arg \min_{\beta} \sum_{x_j \in U} \theta_j * \beta_j (\hat{y}_j - f(x_j))^2 + \lambda \left(\frac{1}{2} \beta^2 - \beta \right) + \mu MMD(S, L \cup Q)$

会发生震荡，需要比较多次或者难以收敛(连续5次选出的样本不变)



观察加入SPL与不加入选择样本的不同

Dataset: ionosphere (200 training samples, 34 attributes)

Origin:

181	26	190	191	30
34	10	127	50	64
101	52	134	170	146

SPL: $\lambda = 1.0$, $pace = 0.15$

26	190	191	181	29
7	69	50	151	147
86	165	143	199	80

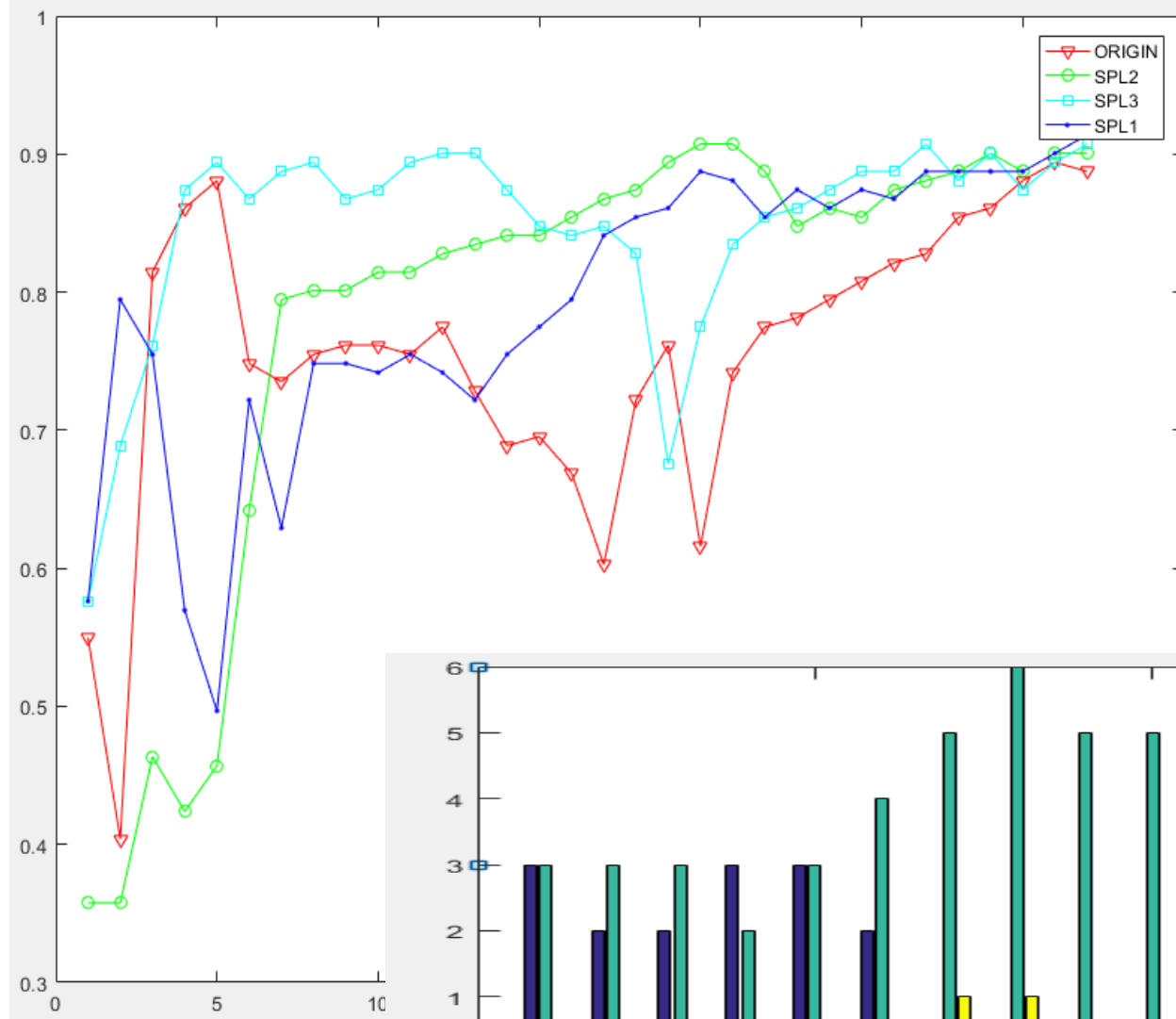
(spl权为0.7, 30号样本spl权为1)

该数据集查到的样本变化比较明显，与iris数据集不一样



真实数据集的准确率变化结果

Dataset: ionosphere (200 training samples, 34 attributes)



SPL1:

$\text{lamda_initial} = 0.2;$

$\text{lamda_pace} = 0.15;$

$\text{lamda_max} = 3.5;$

SPL2:

$\text{lamda_initial} = 0.2;$

$\text{lamda_pace} = 0.1;$

$\text{lamda_max} = 3.5;$

SPL3:

$\text{lamda_initial} = 1.0;$

$\text{lamda_pace} = 0.15;$

$\text{lamda_max} = 3.5;$

