$$\min_{w} \sum_{j=1}^{n_u} w_j \left[l(\hat{y}_j, f(x_j)) + \lambda l(\bar{y}_j, f(x_j)) \right]$$

$$w \in [0,1]^{n_u}$$

$$1^T w = b$$

$$l(y, f(x)) = \max(0, 1 - yf(x))$$

当|f(x)| < 1

$$l\left(\hat{y}_j, f(x_j)\right) + \lambda l\left(\bar{y}_j, f(x_j)\right) = \begin{cases} 1 + |f(x)| + \lambda * (1 - |f(x)|) & \text{当}\bar{y}\text{与模型预测相同} \\ 1 + |f(x)| + \lambda * (1 + |f(x)|) & \text{否则} \end{cases}$$

当|f(x)| >= 1

$$l\left(\hat{y}_j, f(x_j)\right) + \lambda l\left(\bar{y}_j, f(x_j)\right) = \begin{cases} 1 + |f(x)| & \text{当}\bar{y}\text{与模型预测相同} \\ 1 + |f(x)| + \lambda * (1 + |f(x)|) & \text{否则} \end{cases}$$

当$\bar{y}$与模型预测相同时，目标函数的值会更小，倾向于查询两个模型预测一致的样本

当|f(x)| < 1 $l\left(\widehat{y_j}, f\left(x_j\right)\right) + \lambda l\left(\overline{y_j}, f\left(x_j\right)\right) = \begin{cases} 1 + |f(x)| + \lambda * (1 - |f(x)|) & \text{当}\bar{y}\text{与模型预测相同} \\ 1 + |f(x)| + \lambda * (1 + |f(x)|) & \text{否则} \end{cases}$

$\lambda < 1$	$\|f(x)\| + const$	等价于uncertainty
$\lambda = 1$	$const$	Margin内随机选
$\lambda > 1$	$-\|f(x)\| + const$	选靠近margin的点

当|f(x)| < 1

$$l\left(\widehat{y_j}, f\left(x_j\right)\right) + \lambda l\left(\overline{y_j}, f\left(x_j\right)\right) = \begin{cases} 1 + |f(x)| + \lambda * (1 - |f(x)|) & \text{当}\bar{y}\text{与模型预测相同} \\ 1 + |f(x)| + \lambda * (1 + |f(x)|) & \text{否则} \end{cases}$$

当|f(x)| >= 1

$$l\left(\widehat{y_j}, f\left(x_j\right)\right) + \lambda l\left(\overline{y_j}, f\left(x_j\right)\right) = \begin{cases} 1 + |f(x)| & \text{当}\bar{y}\text{与模型预测相同} \\ 1 + |f(x)| + \lambda * (1 + |f(x)|) & \text{否则} \end{cases}$$

$\lambda < 1$

[1,2]
[1,4]
[2,+]
[2,+]

完全查询margin内的点
(倾向于一致的样本)

$\lambda = 1$

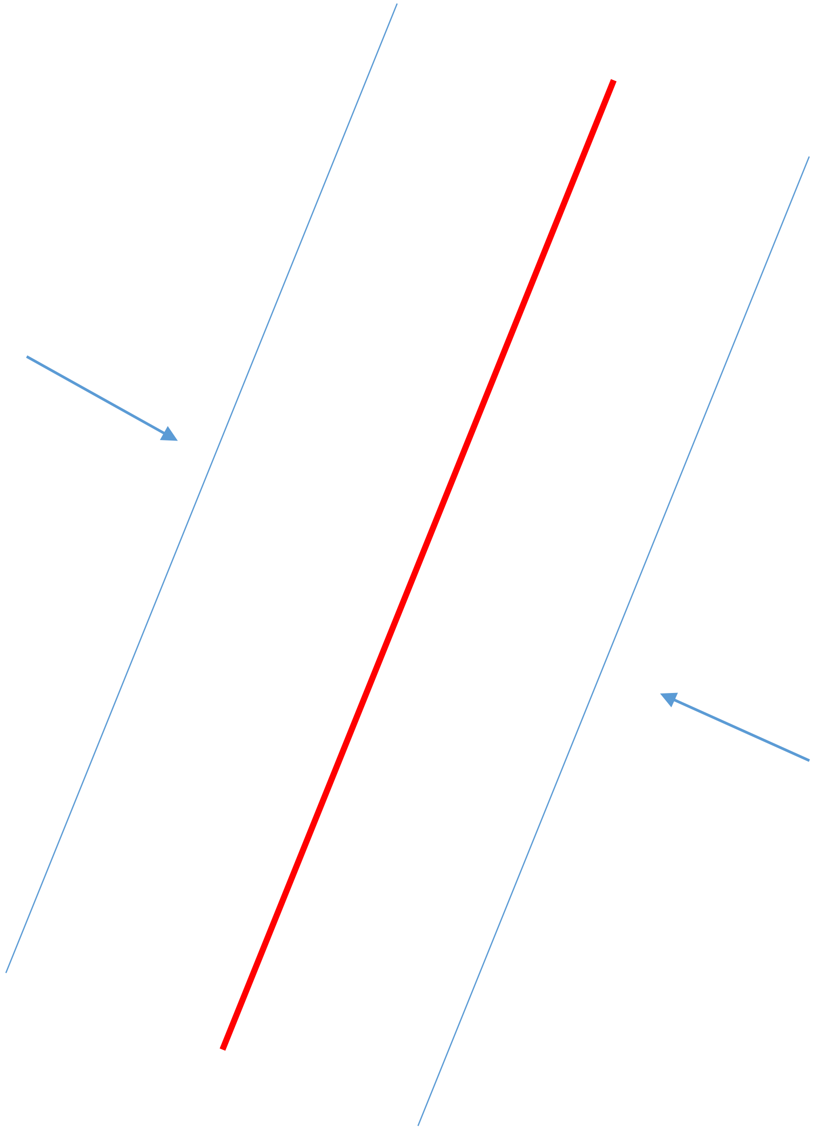
2
[2,4]
1+ |f(x)|>2
2+ 2|f(x)|>4

完全查询margin内的点
(在一致的点内随机)

$\lambda > 1$

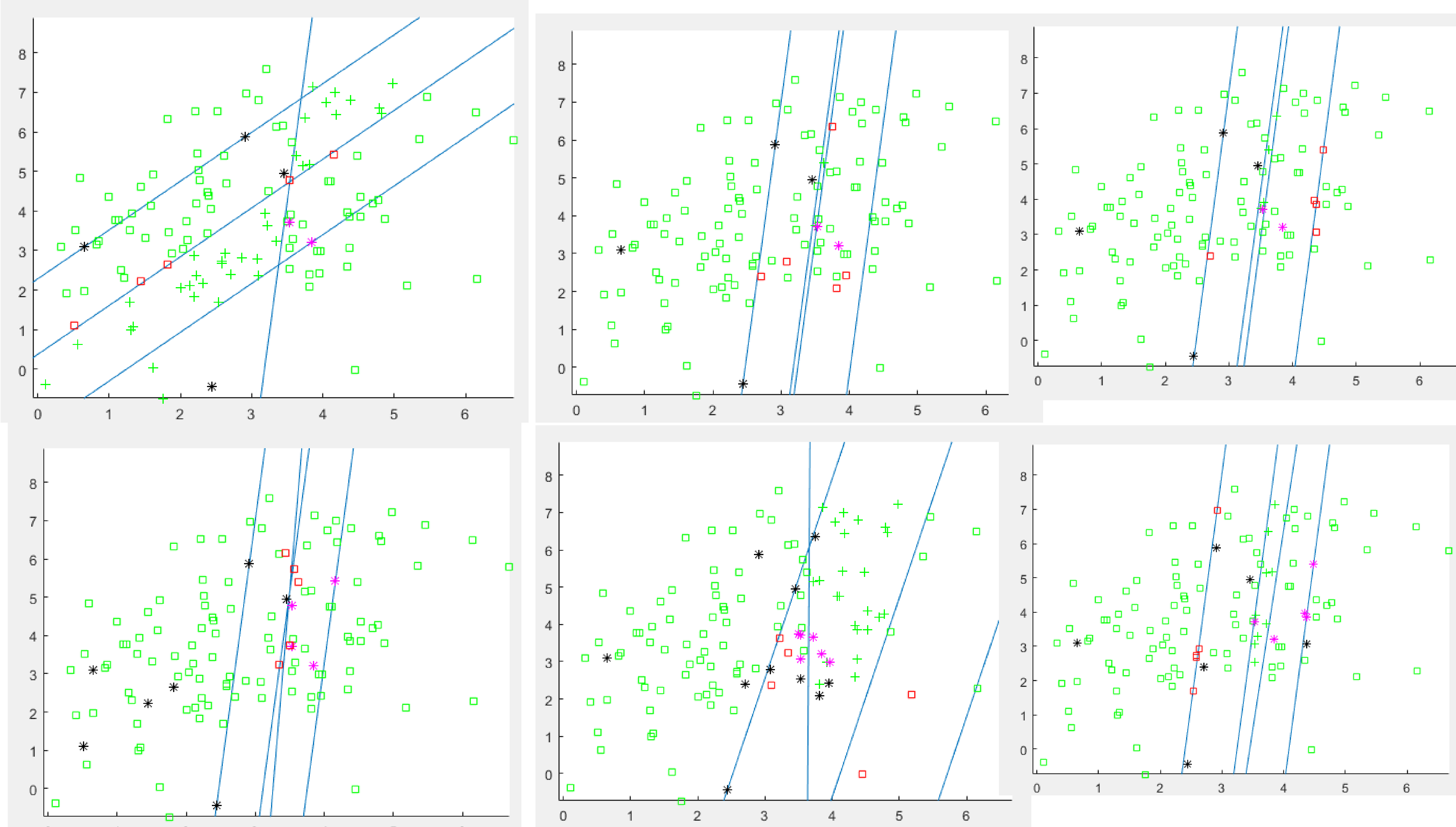
[2, +]
(2,+]
[2,+]
(4,+]

靠近margin上的点



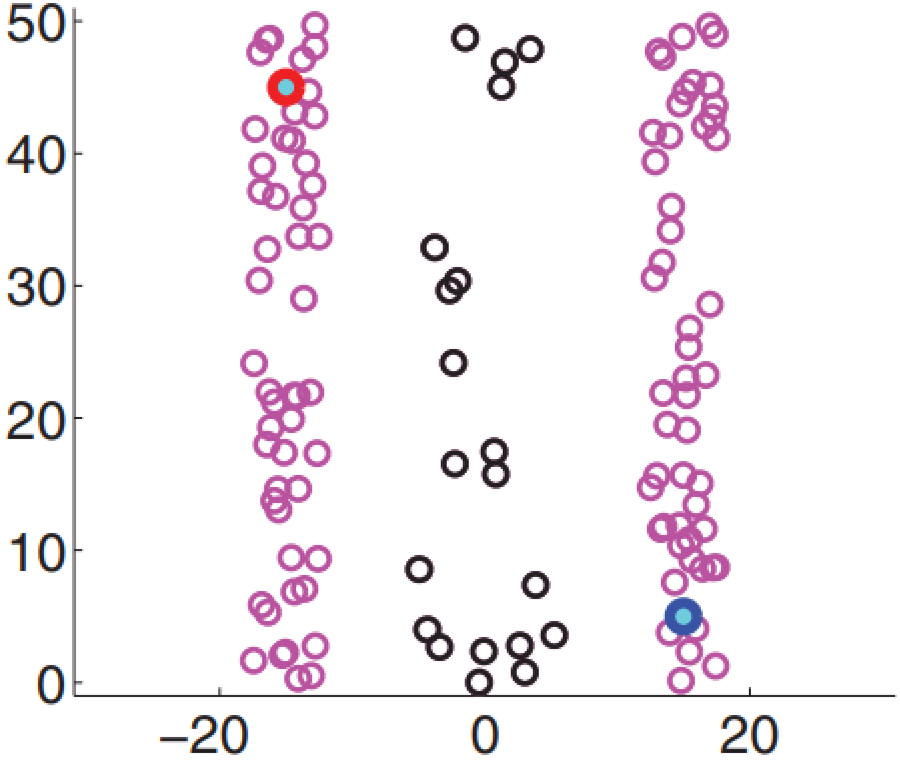
当存在margin内的样本时，仅选择其中的样本：选择策略与λ相关

$$\begin{cases}\lambda < 1 & \text{等价于uncertainty，倾向于一致的点} \\ \lambda = 1 & \text{Margin内一致的点随机选} \\ \lambda > 1 & \text{选择靠近margin的点}\end{cases}$$



Large-Scale Adaptive Semi-Supervised Learning via Unified Inductive and Transductive Model

KDD 14 De Wang , Feiping Nie, Heng Huang



boundary points will blur the clear distribution of the whole data, and are very noisy for learning。

$$\min_{W,b,Y} \left\| X_l^T W + \mathbf{1}_{nl} b^T - Y_l \right\|_F^2 + \sum_{i=1}^{n} \sum_{k=1}^{c} y_{ik}^r \left\| x_i^T W + b^T - t_k^T \right\|_F^2$$

$$s.t.\ \ \forall i, y_{ik} \in [0,1], \sum_{k=1}^{c} y_{ik} = 1 \qquad (1)$$

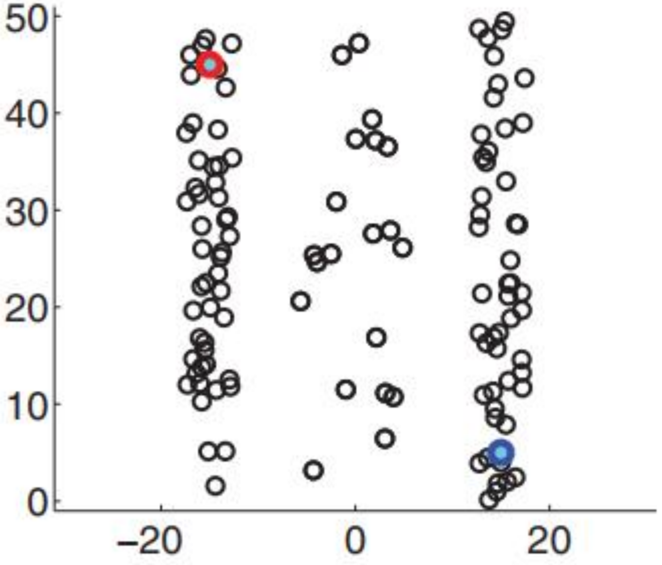
those clear classified points still have large weights in total and contribute a lot to the second term in the objective function

$y_{i1} = 0.9$ $y_{i2} = 0.1$ $y_{i1}^2 = 0.81$ $y_{i2}^2 = 0.01$

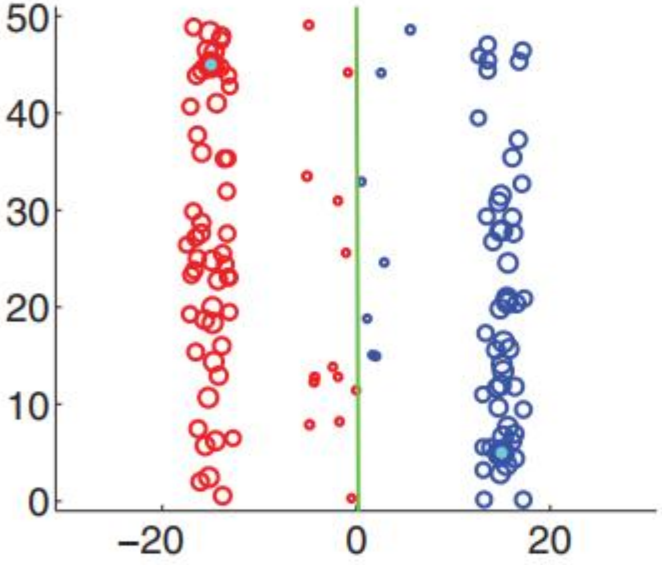
For boundary points, however, y_{i1} and y_{i2} would be more likely equal

$y_{i1} = 0.5$ $y_{i2} = 0.5$ $y_{i1}^2 = 0.25$ $y_{i2}^2 = 0.25$

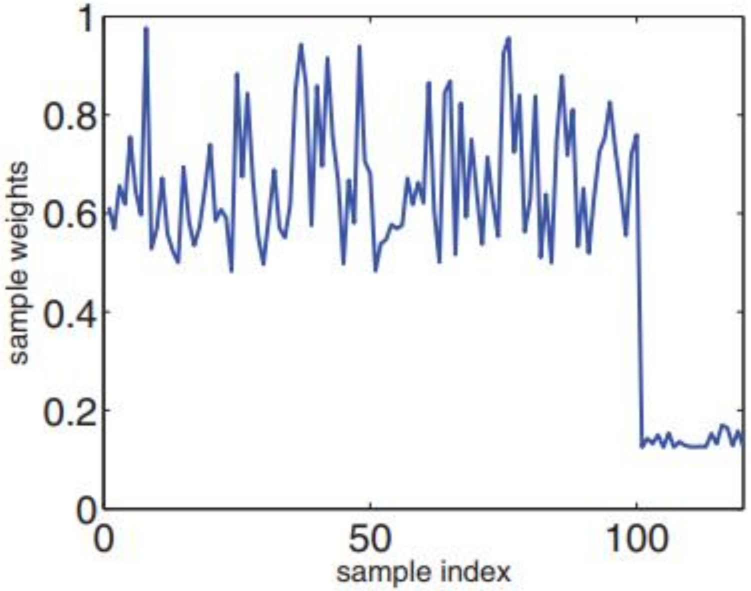
If r is too large, the y_{ik}^r will tend to be zero, and the unsupervised information is lost. So r can not be too large. That's why r is suggested to vary in [1,2] (Consider the class number is often larger



(a) Original toy data



(b) Toy data after classification using our model



(c) Sample weights

Table 1: Data Sets Descriptions

	# sample	# feature	# class
AR	840	768	120
YALE-B	2414	1024	38
MSRC-V1	1799	1024	12
CMU-PIE	3329	1024	68
FERET	1400	1296	200
ORL	400	644	40

Table 3: Classification accuracy and standard deviation for running different semi-supervised learning methods 20 times on YALE-B Data Set.

YALE-B	kl=1		kl=3		kl=5	
	Unlabeled	Testing	Unlabeled	Testing	Unlabeled	Testing
LGC	42.47±2.11	NA	60.42±2.42	NA	68.16±1.48	NA
RW	43.98±2.43	NA	61.37±2.38	NA	68.69±1.42	NA
GFHF	43.75±1.96	NA	63.05±1.71	NA	70.32±2.12	NA
LapReg	51.36±2.88	50.88±3.11	86.09±1.73	86.51±1.92	94.90±0.82	95.02±1.07
FME	51.24±2.01	50.90±3.17	85.70±2.55	85.54±2.92	94.61±2.03	94.83±2.11
ASL	**59.35±2.91**	**58.98±8.08**	**96.06±1.86**	**96.13±1.92**	**99.14±0.54**	**99.00±0.85**

Table 4: Classification accuracy and standard deviation for running different semi-supervised learning methods 20 times on MSRC-V1 Data Set.

MSRC-V1	kl=1		kl=3		kl=5	
	Unlabeled	Testing	Unlabeled	Testing	Unlabeled	Testing
LGC	62.38±3.46	NA	88.34±5.38	NA	95.19±3.58	NA
RW	60.05±3.78	NA	88.98±5.29	NA	96.88±2.94	NA
GFHF	55.73±6.01	NA	89.73±4.39	NA	97.07±2.72	NA
LapReg	81.91±4.44	81.83±4.65	97.54±2.08	97.55±1.90	98.69±1.69	98.61±1.61
FME	80.72±4.34	80.45±4.50	97.17±2.64	97.28±2.57	99.29±0.88	99.22±0.98
ASL	**82.55±3.64**	**82.29±3.74**	**98.39±1.92**	**98.45±1.82**	**99.36±1.25**	**99.42±1.15**

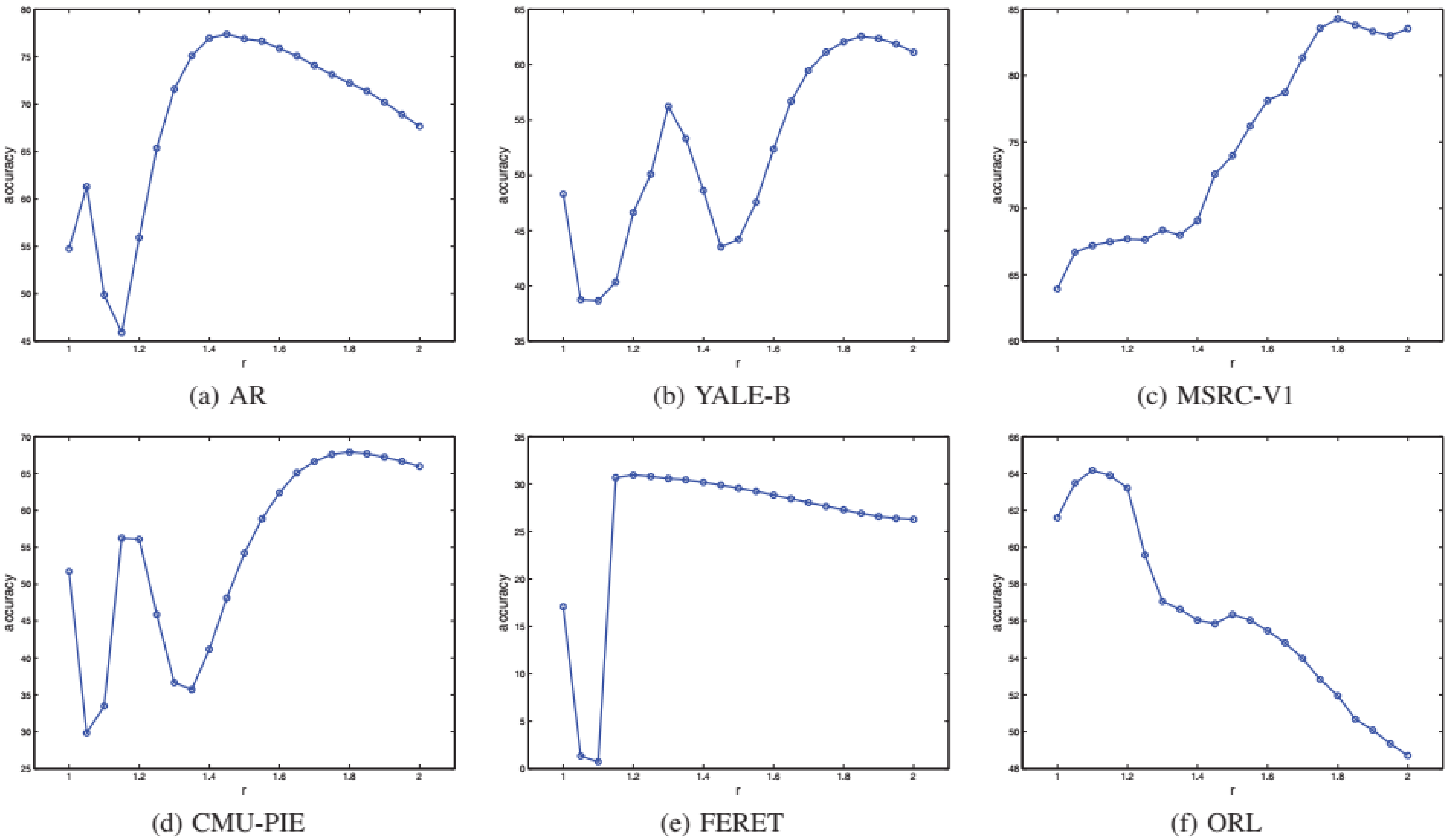
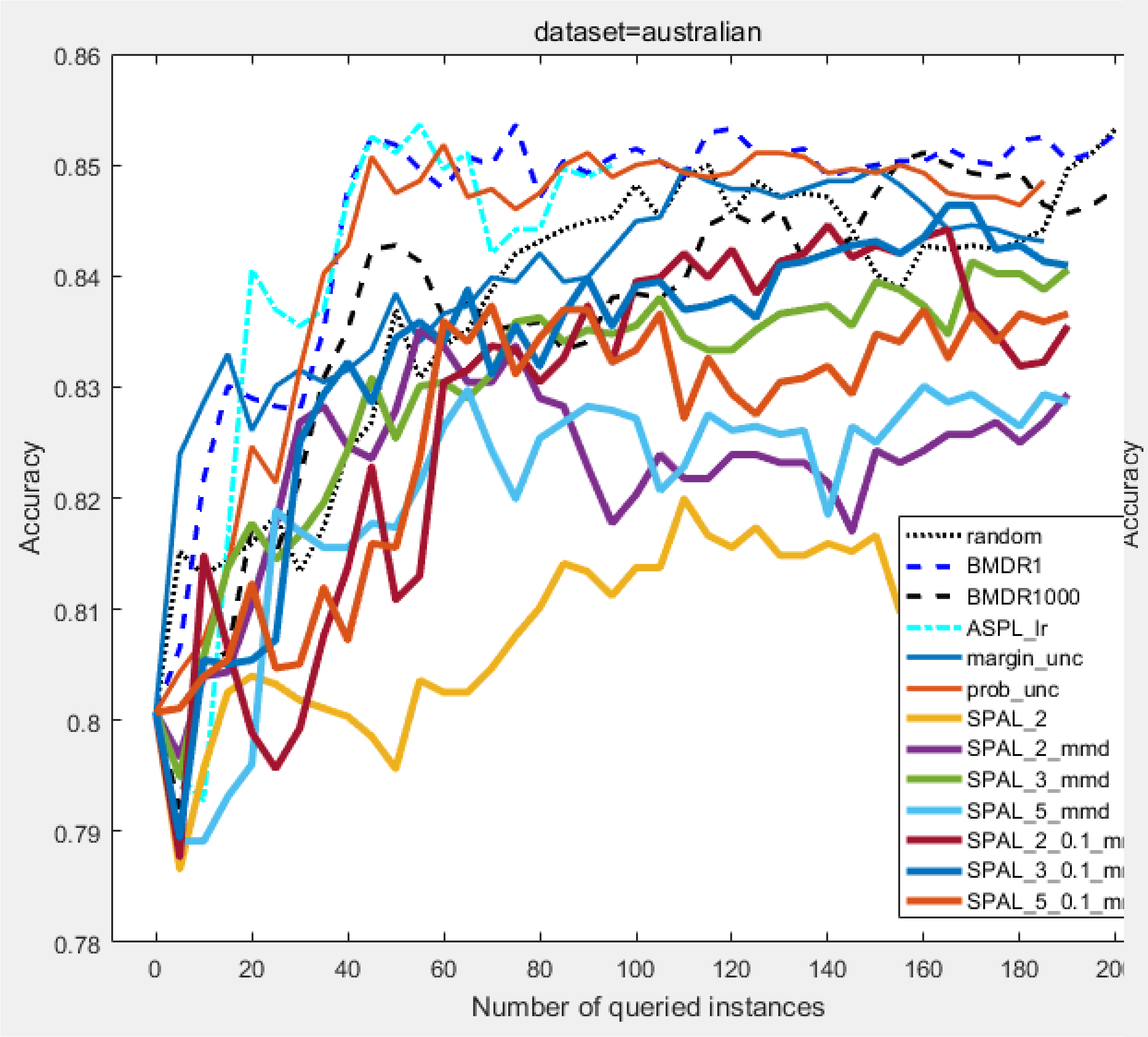
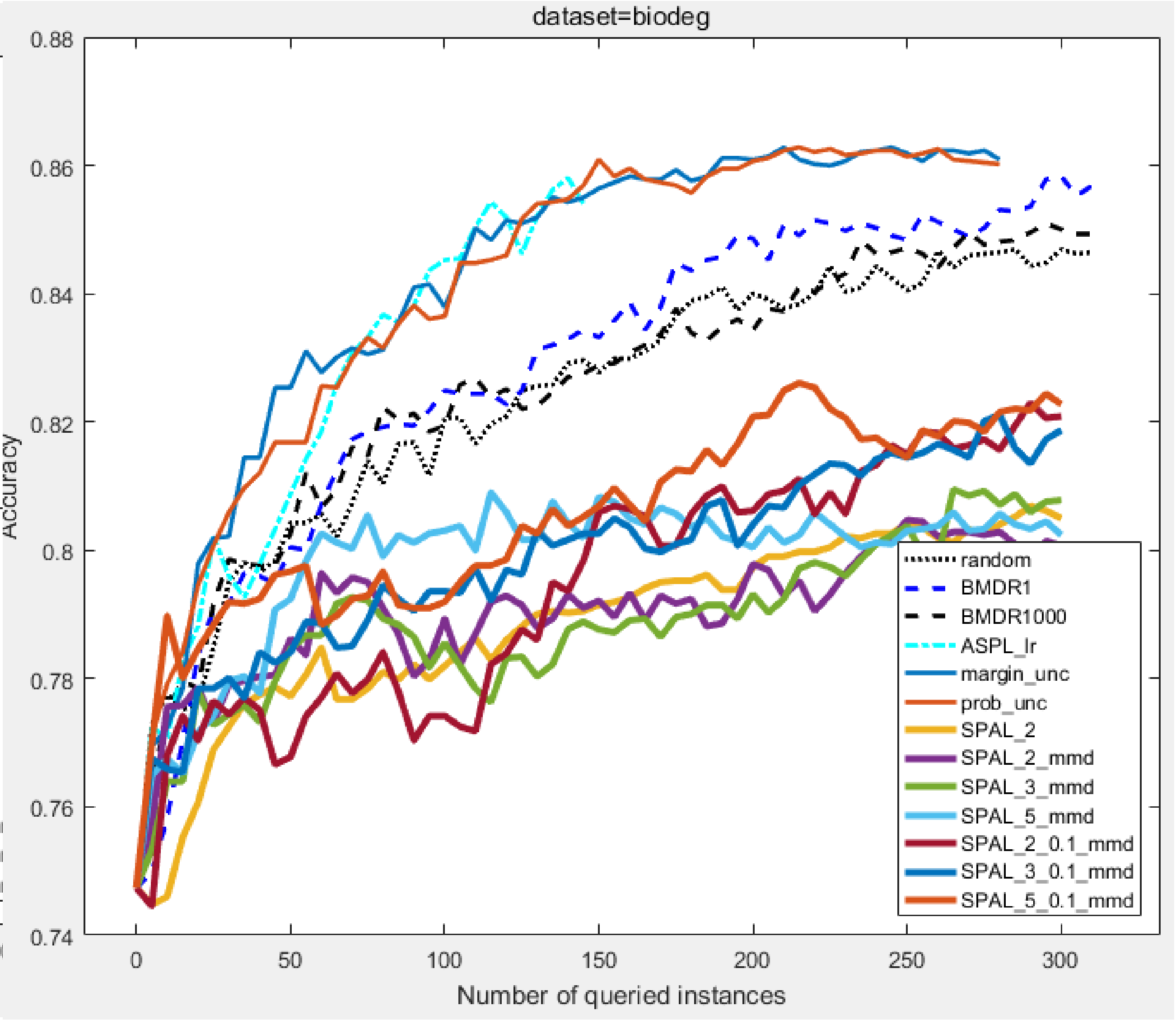
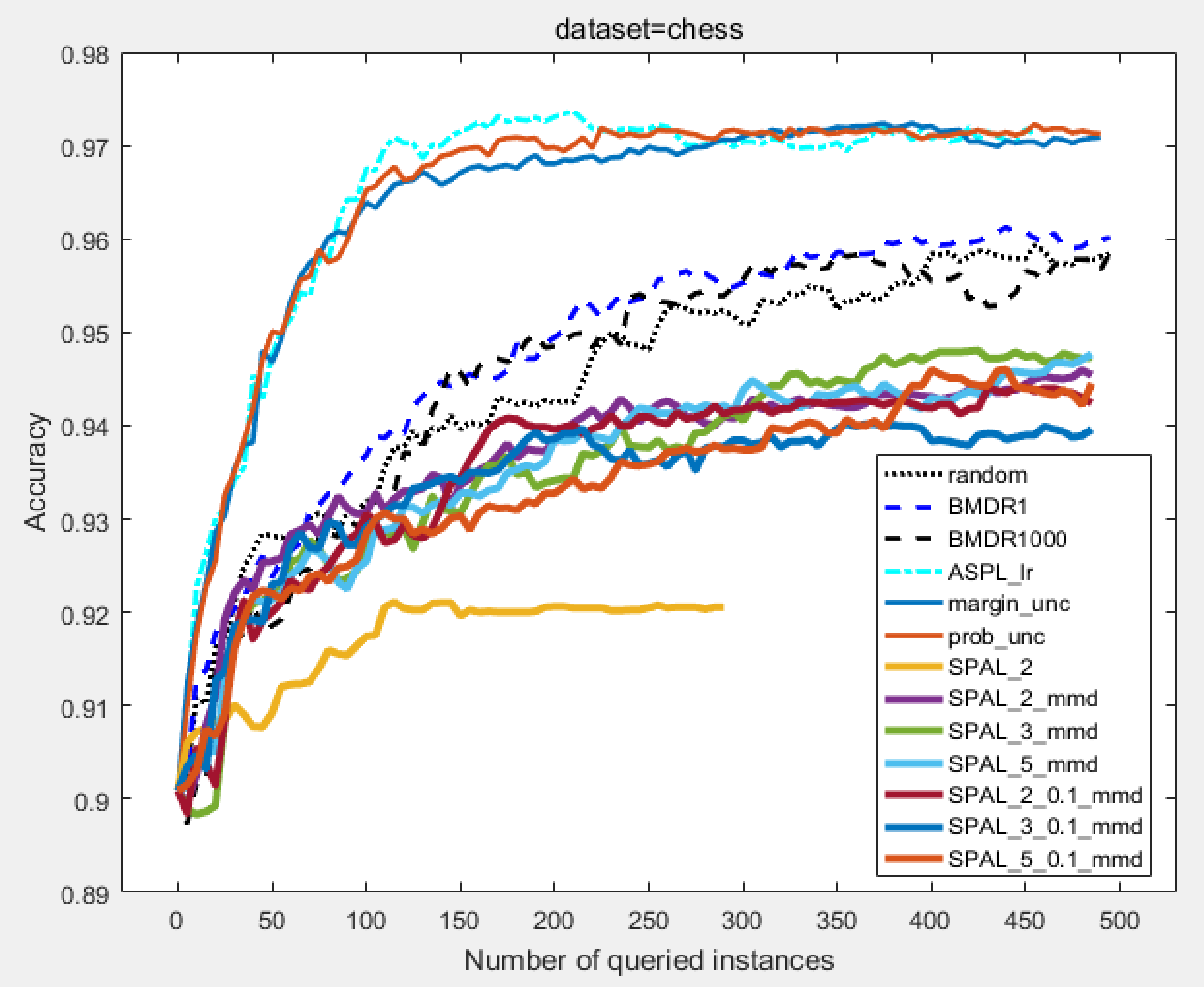
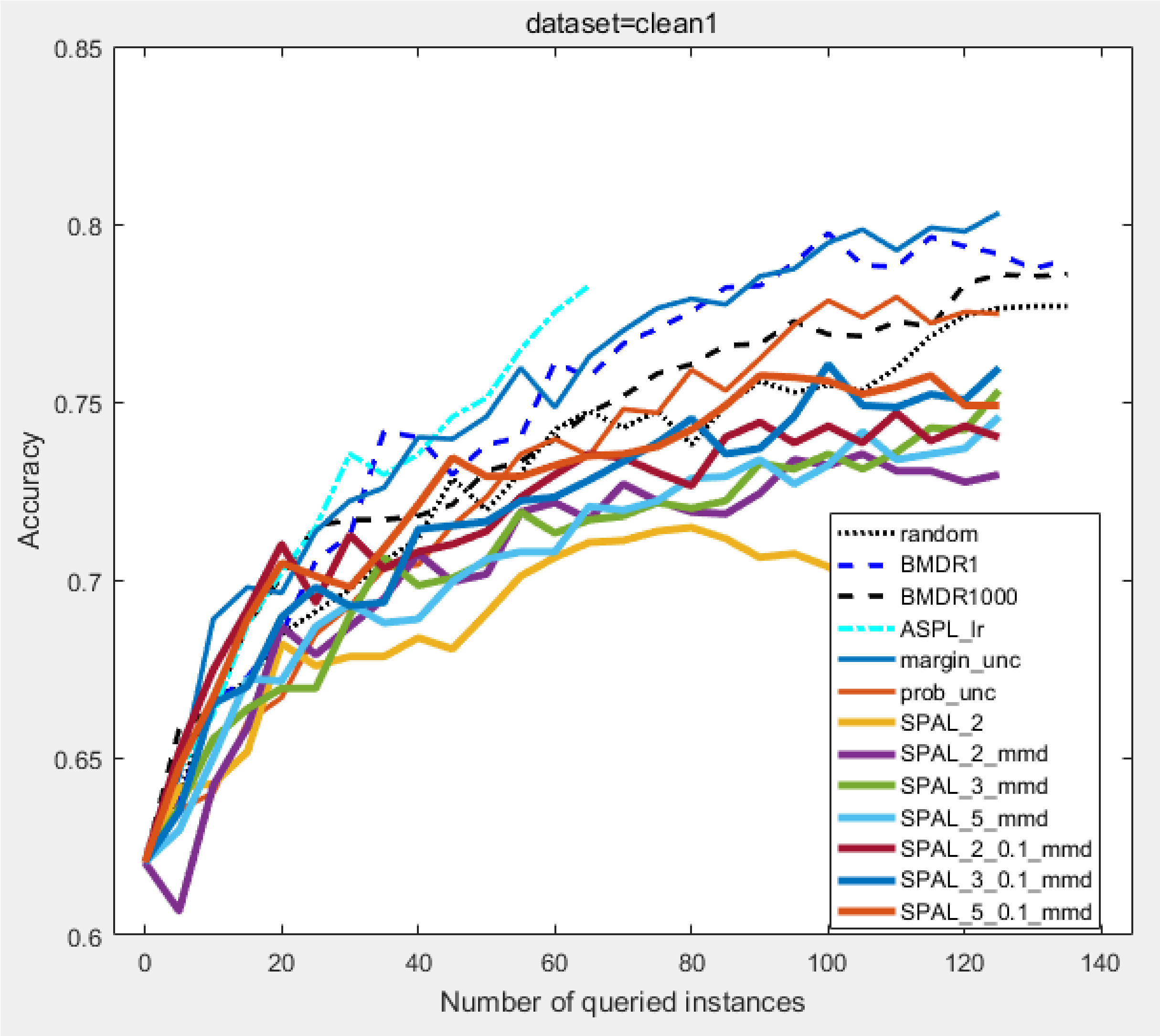


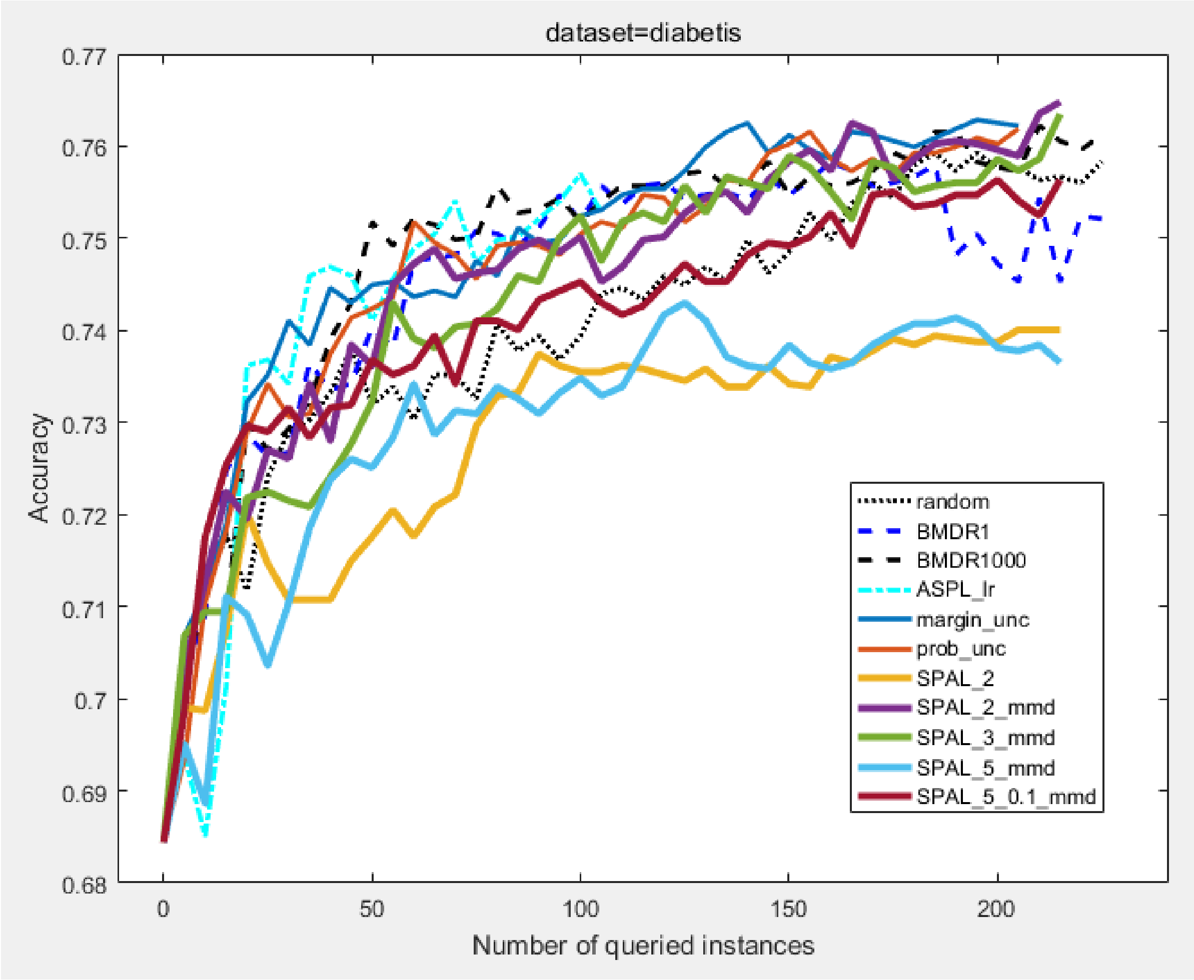
Figure 5: The classification accuracy using different r values on the six data sets. One labeled data point is used from each class.

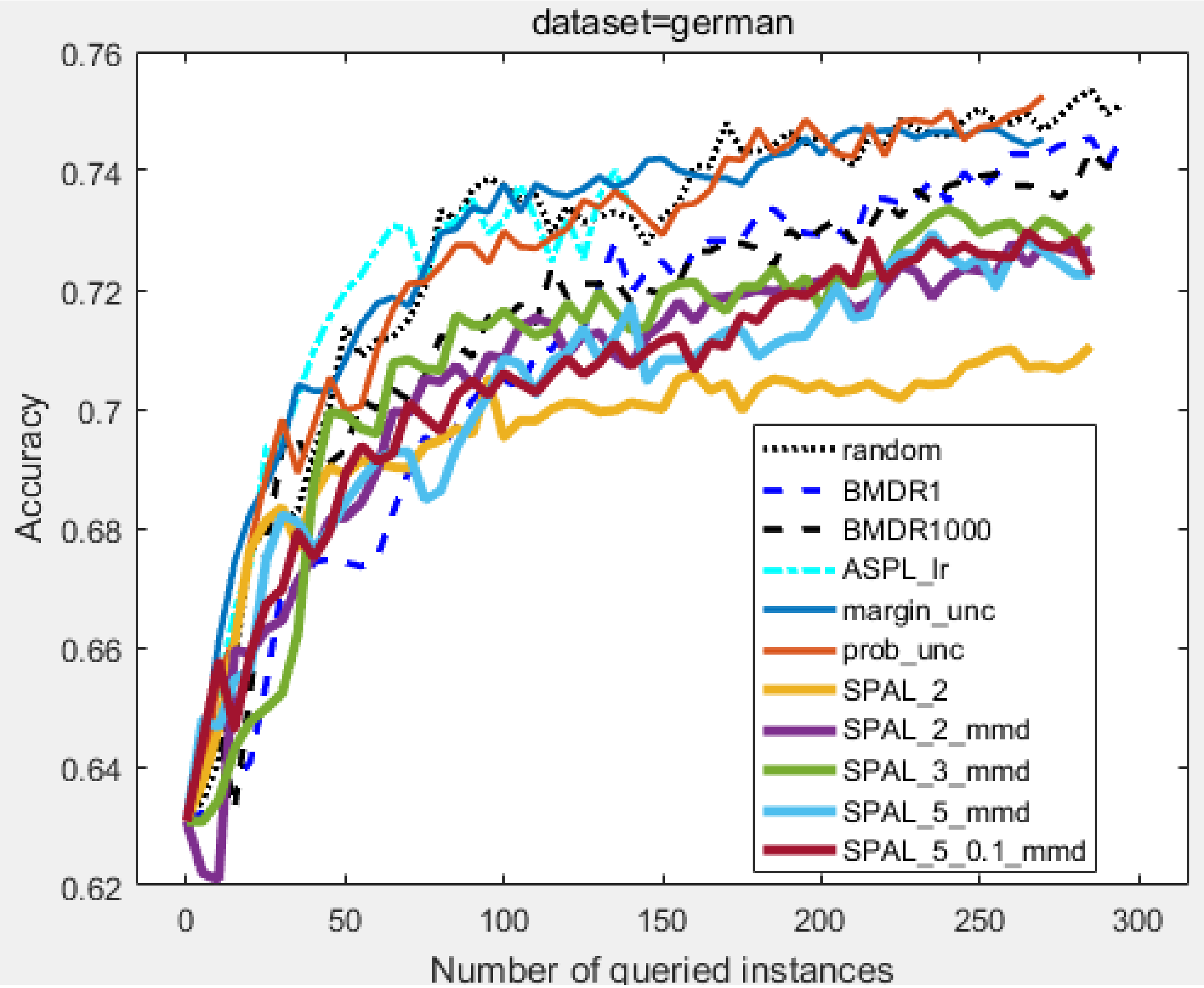

dataset=australian
Accuracy
Number of queried instances
random
BMDR1
BMDR1000
ASPL_lr
margin_unc
prob_unc
SPAL_2
SPAL_2_mmd
SPAL_3_mmd
SPAL_5_mmd


dataset=biodeg
Accuracy
Number of queried instances
random
BMDR1
BMDR1000
ASPL_lr
margin_unc
prob_unc
SPAL_2
SPAL_2_mmd
SPAL_3_mmd
SPAL_5_mmd
SPAL_2_0.1_mmd
SPAL_3_0.1_mmd
SPAL_5_0.1_mmd


dataset=chess
Accuracy
Number of queried instances
0.98
0.97
0.96
0.95
0.94
0.93
0.92
0.91
0.9
0.89
0
50
100
150
200
250
300
350
400
450
500
random
BMDR1
BMDR1000
ASPL_lr
margin_unc
prob_unc
SPAL_2
SPAL_2_mmd
SPAL_3_mmd
SPAL_5_mmd
SPAL_2_0.1_mmd
SPAL_3_0.1_mmd
SPAL_5_0.1_mmd


dataset=clean1
Accuracy
Number of queried instances
0.85
0.8
0.75
0.7
0.65
0.6
0
20
40
60
80
100
120
140
random
BMDR1
BMDR1000
ASPL_lr
margin_unc
prob_unc
SPAL_2
SPAL_2_mmd
SPAL_3_mmd
SPAL_5_mmd
SPAL_2_0.1_mmd
SPAL_3_0.1_mmd
SPAL_5_0.1_mmd


dataset=diabetis
Accuracy
Number of queried instances
0.77
0.76
0.75
0.74
0.73
0.72
0.71
0.7
0.69
0.68
0
50
100
150
200
random
BMDR1
BMDR1000
ASPL_lr
margin_unc
prob_unc
SPAL_2
SPAL_2_mmd
SPAL_3_mmd
SPAL_5_mmd
SPAL_5_0.1_mmd


dataset=german
Accuracy
Number of queried instances
0.76
0.74
0.72
0.7
0.68
0.66
0.64
0.62
0
50
100
150
200
250
300
random
BMDR1
BMDR1000
ASPL_lr
margin_unc
prob_unc
SPAL_2
SPAL_2_mmd
SPAL_3_mmd
SPAL_5_mmd
SPAL_5_0.1_mmd