极端支持向量机ESVM及增量学习与样本选择

何清　研究员
中国科学院计算技术研究所

目录

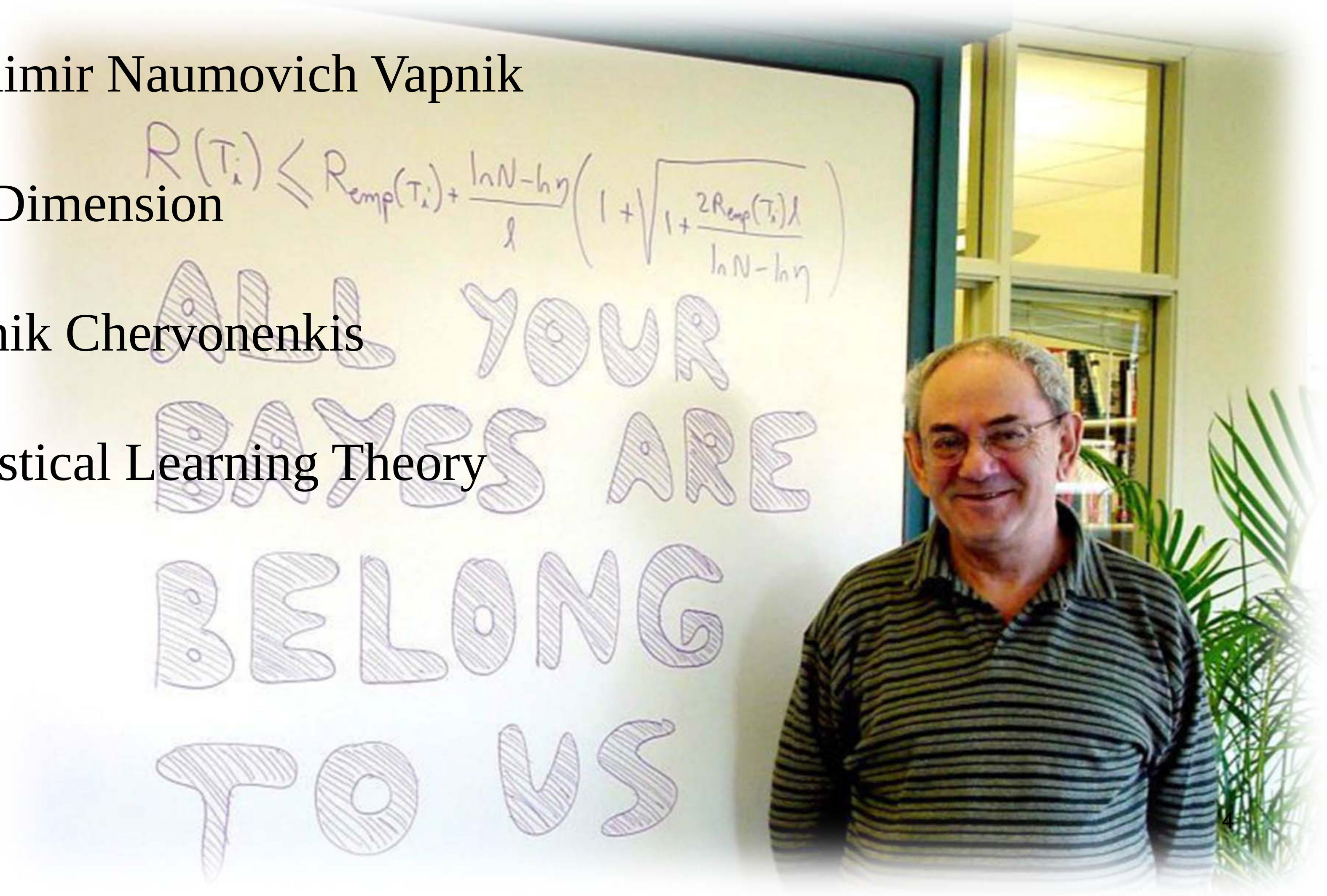
- 引言
- 非线性PSVM增量学习算法
- PSVM样本选择技术
- 极端支持向量：ESVM

目录

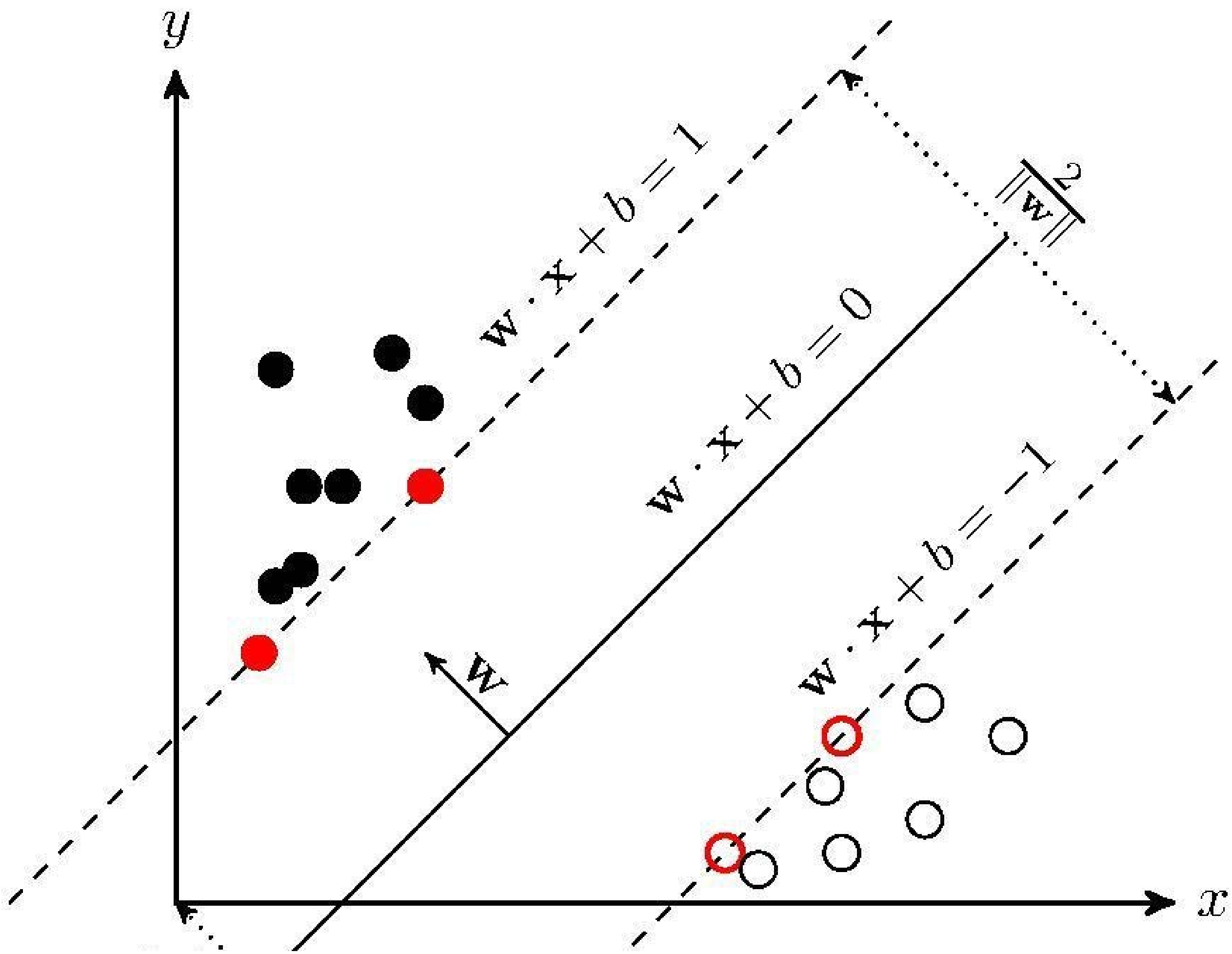
- 引言
- 非线性PSVM增量学习算法
- PSVM样本选择技术
- 极端支持向量：ESVM

预备知识-SVM

- Vladimir Naumovich Vapnik
- VC Dimension
- Vapnik Chervonenkis
- Statistical Learning Theory



预备知识-SVM



y
x
$\mathbf{w} \cdot \mathbf{x} + b = 1$
$\mathbf{w} \cdot \mathbf{x} + b = 0$
$\mathbf{w} \cdot \mathbf{x} + b = -1$
$\frac{2}{\|\mathbf{w}\|}$
$\mathbf{w}$

预备知识-PSVM

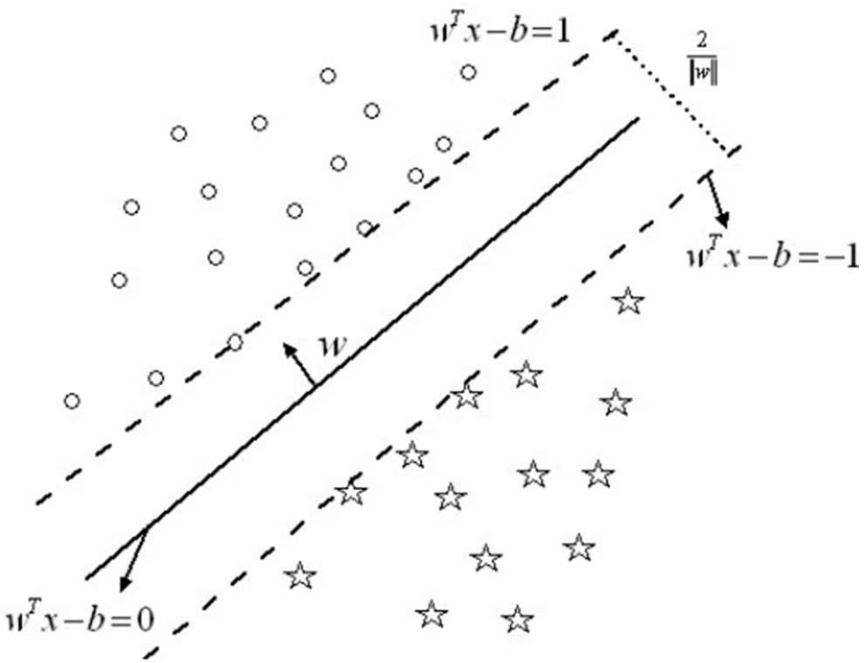
Proximal support vector machine classifiers

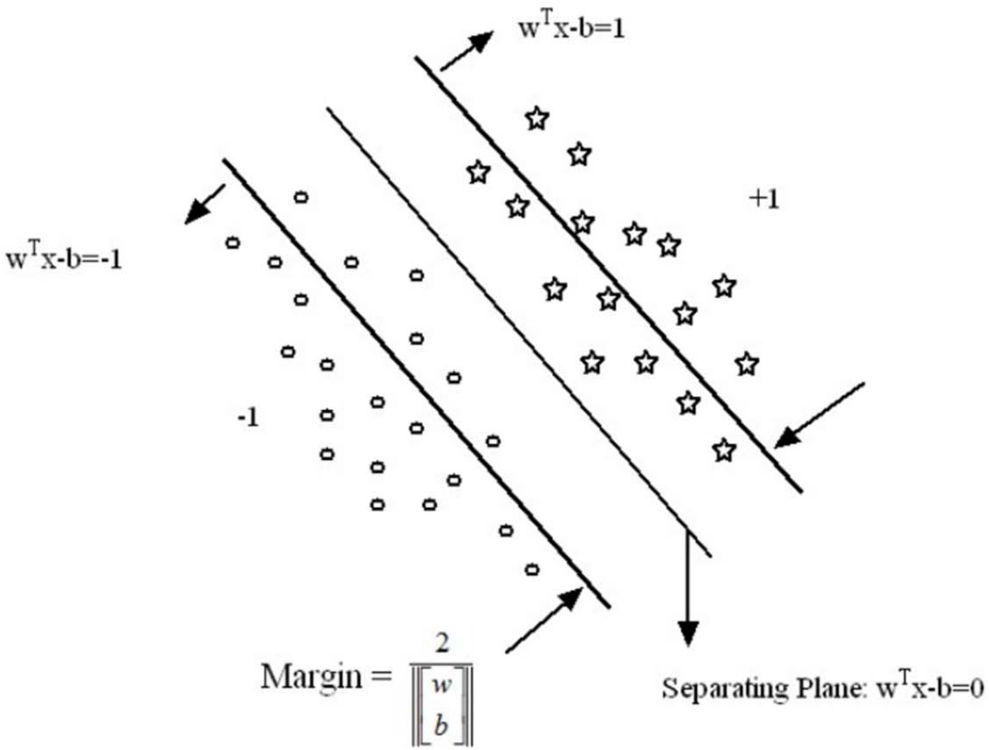
https://dl.acm.org › citation - 翻译此页

作者：G Fung - 2001 - 被引用次数：1190 - 相关文章

2001年8月26日 - Computational results on publicly available datasets indicate that the proposed proximal SVM classifier has comparable test set correctness to ...

Abstract · References · Cited By · Publication





研究工作简介

- 设计了非线性PSVM的增量学习算法
- 设计了PSVM的样本选择技术
- 提出了一种新的SVM分类模型

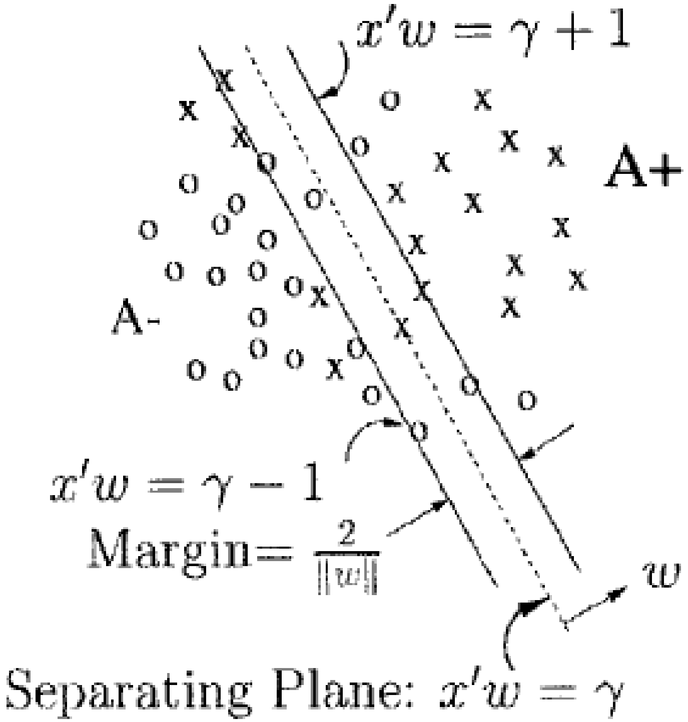
目录

- 引言
- 非线性PSVM增量学习算法
- PSVM样本选择技术
- 极端支持向量：ESVM

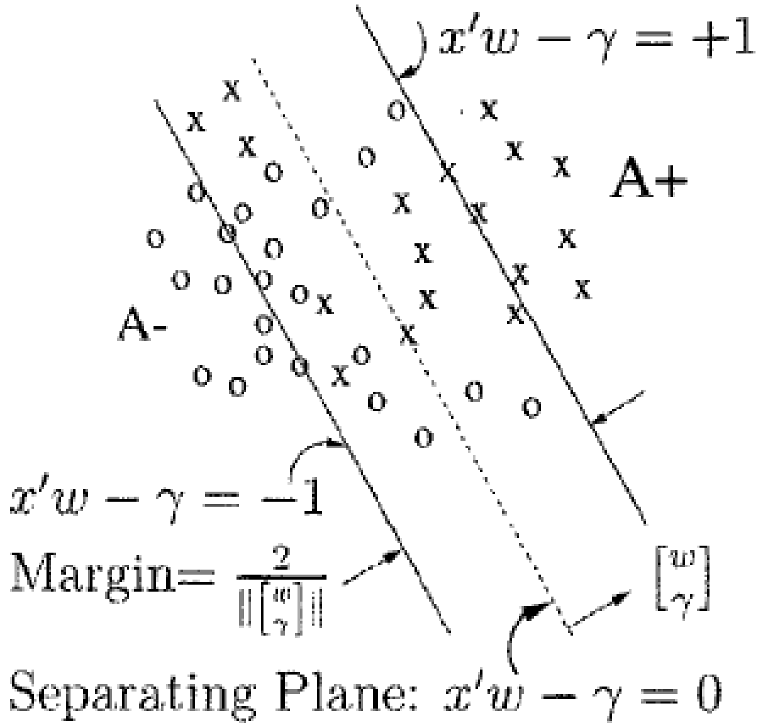
非线性PSVM增量学习算法

- 易于增量学习的Nonlinear PSVM分类器;
- 利用分块矩阵求逆公式，设计了PSVM增量学习算法；
- 理论及实验证明：在线学习过程中采用该算法不仅可以得到与批量学习相同的分类器、正确率；而且解决了重复学习的问题，可以显著地缩短训练时间。

非线性PSVM增量学习算法



SVM



Proximal SVM

非线性PSVM增量学习算法

为了叙述的方便，首先把这里将用到的数学表示符号声明如下：所有的向量默认均为列向量，上标'表示转置，向量 x 和 y 在空间中的点积表示为：$x' y$，向量 x 的 2-norm 表示为 $\|x\| = \sqrt{x' \cdot x}$；这里我们利用矩阵 $A[m \times n]$ 表示 n 维空间 R^n 中的 m 个训练样本；对角线元素为 ± 1 的对角矩阵 $D[m \times n]$ 的对角线上的元素声明了 m 个训练样本的类别是 $+1$ 或 -1；对于矩阵 $A \in R^{m \times n}$ 和 $B \in R^{n \times l}$，核函数 $k(A, B)$ 将 $R^{m \times n}, R^{n \times l}$ 映射到 $R^{m \times l}$；默认情况下我们将使用下面的 Gaussian 核：

$$\left(K(A, B)\right)_{i,j} = \varepsilon^{-\mu\|A_i' - B_{\cdot j}\|^2}, i = 1..m, j = 1..l$$

其中 μ 是正常量，ε 表示自然对数的底；e 为元素为 1 的任意维向量（维度根据上下文确定）；w, r 分别为分类超平面的方向系数和偏置，y 为松弛向量，参数 $v \geq 0$ 用于控制模型复杂度和准确率之间的平衡；I 表示单位向量。

非线性PSVM增量学习算法

$$\min_{(w,\gamma,y)\in R^{n+1+m}} \quad \nu\tfrac{1}{2}\|y\|^2+\tfrac{1}{2}(w'w+\gamma^2)$$
$$\text{s.t.} \quad D(Aw-e\gamma)+y = e$$

Nonlinear PSVM

$$\min_{(u,r,y)\in R^{m+1+m}} v\frac{1}{2}\|y\|^2+\frac{1}{2}(u'u+r^2),$$
$$s.t.\quad D(K(A,A')Du-er)+y=e.$$

$$\min_{(w,r,y)\in R^{m+1+n}} v\frac{1}{2}\|y\|^2+\frac{1}{2}\left(w'w+r^2\right),$$
$$s.t.\quad D\left(\phi(A)\mathrm{w}-er\right)+y=e.$$

新NPSVM模型

NPSVM的解

$$Ds=(\frac{I}{v}+KK'+ee')^{-1}De$$

新NPSVM模型的解

$$Du=(I/v+\phi(A)\phi(A)'+ee')^{-1}De$$
$$=(I/v+K+ee')^{-1}De$$

非线性PSVM增量学习算法

$$Du = (I/v + \phi(A)\phi(A)' + ee')^{-1} De$$
$$= (I/v + K + ee')^{-1} De$$

新NPSVM模型的解

$$A_{11}^{-1} = \left(I/v + K\left(A_1, A_1'\right) + ee'\right)^{-1}$$

历史训练结果

$$A_n = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

新数据

where $A_{12} = \left(K\left(A_1, A_2'\right) + ee'\right), A_{22} = \left(I/V + K\left(A_2, A_2'\right) + ee'\right)$ and $A_{21} = A_{12}'$.

$$A_n^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} + X & Y \\ Y' & T \end{bmatrix},$$

$$T = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}, Y = -A_{11}^{-1}A_{12}T, X = -YA_{12}'A_{11}^{-1}.$$

历史训练结果的更新

非线性PSVM增量学习算法

	Data Set $m \times n$	Ionosphere 351×34	Bupa Liver 345×6	Tic-Tac-Toe 958×9
NPSVM	Recall Test Time(Sec.)	99.43% 93.17% 0.16	80.23% 66.87% 0.16	100% 99.06% 2.32
NNPSVM	Recall Test Time(Sec.)	99.94% 94.80% 0.14	84.73% 67.27% 0.13	100% 99.06% 1.56

	Data Set $m \times n$	Ionosphere 351×34	Bupa Liver 345×6	Tic-Tac-Toe 958×9
INPSVM	Recall Time(Sec.)	99.72% 0.21	83.48% 0.21	100.00% 3.75
NNPSVM	Recall Time(Sec.)	99.72% 0.33	83.48% 0.33	100.00% 5.03

非线性PSVM增量学习算法

- Qiuge Liu, Qing He, and Zhongzhi Shi. Incremental Nonlinear Proximal Support Vector Machine, in D. Liu et al. (Eds.): ISNN 2007, LNCS 4493, Part III, pp. 336–341, 2007(EI已收录).

目录

- 引言
- 非线性PSVM增量学习算法
- PSVM样本选择技术
- 极端支持向量：ESVM

PSVM样本选择技术

矩阵大小！？

$$A_{\pi}^{-1}=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1}=\begin{bmatrix} A_{11}^{-1}+X & Y \\ Y' & T \end{bmatrix},$$

样本选择

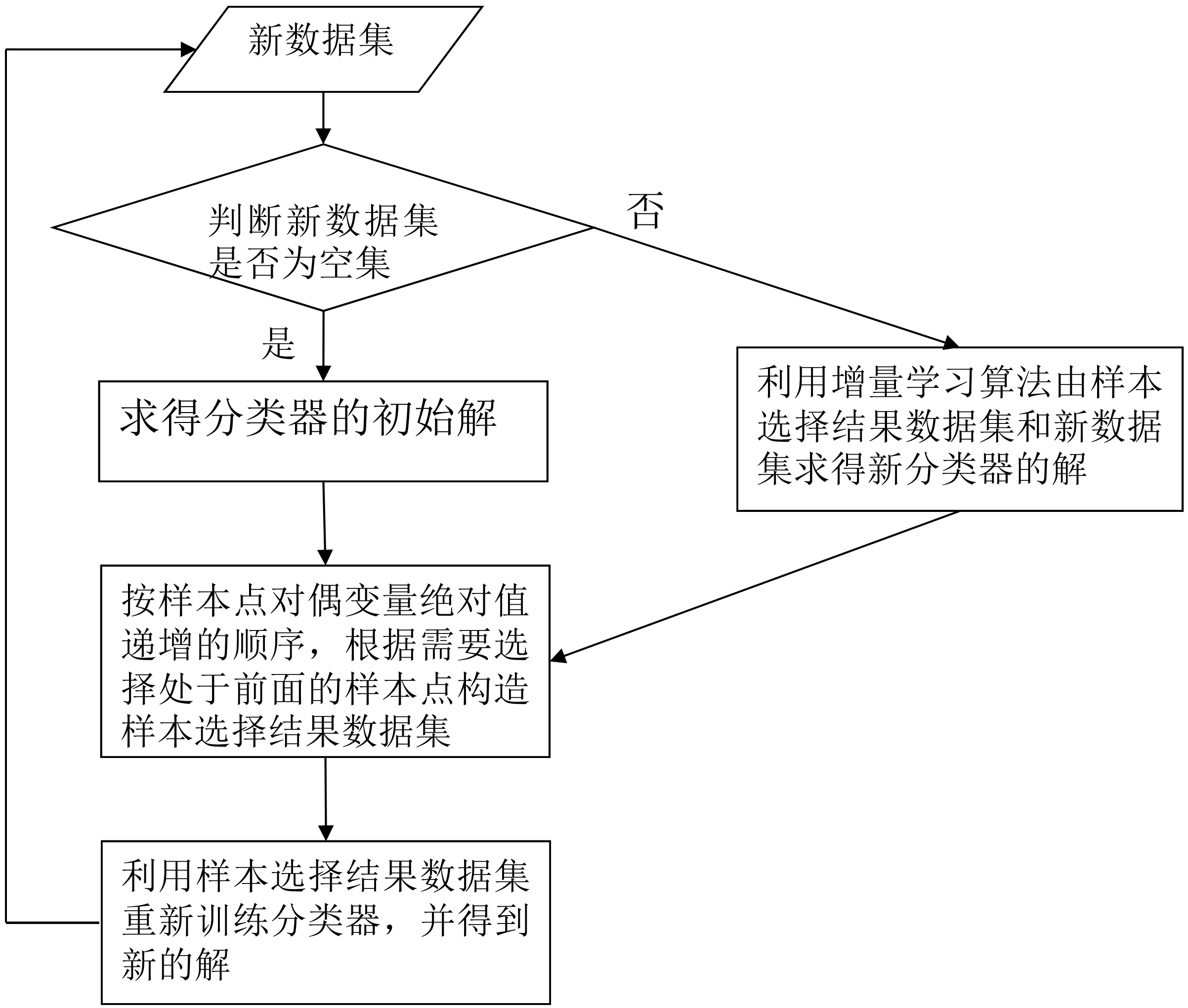
PSVM样本选择技术

- 我们设计了根据训练结果对样本进行选择的技术
  - 历史数据
  - 新数据
  - 非线性分类器
- 与增量学习算法结合，大数据集的增量学习框架

PSVM样本选择技术

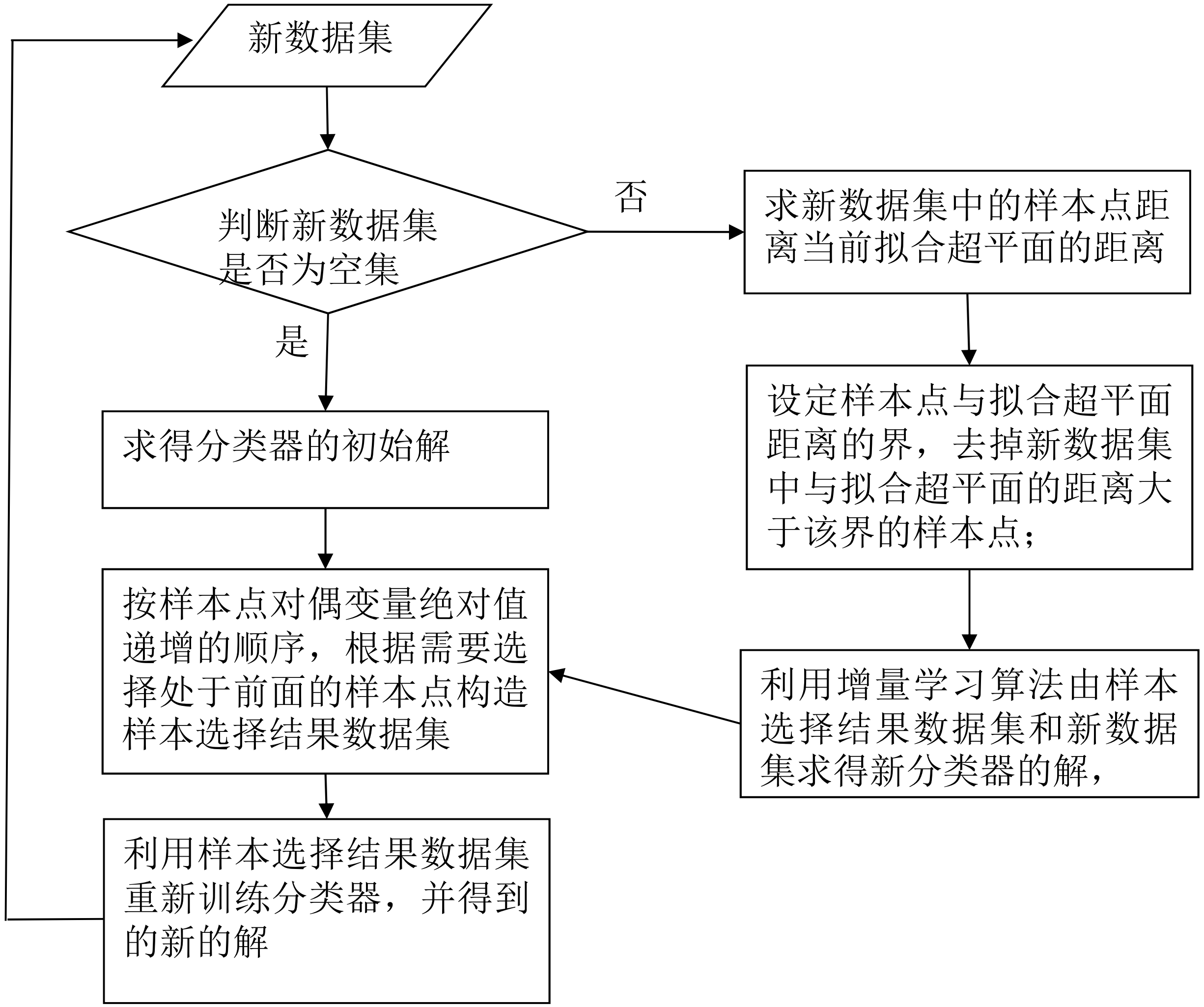
- SVM：保留对偶变量值为非零的点——支持向量
- 实验结果表明：距离拟合超平面较近的样本点具有较强的描述能力，利用这些样本点重新训练可得到与原分类器相似的结果
- PSVM：ε*支持向量* 概念——对偶变量绝对值小于ε的样本点

PSVM样本选择技术



对历史数据的选择

PSVM样本选择技术



添加了对新样本选择的过程

PSVM样本选择技术

- 非线性分类器情形：由于非线性分类器难以描述，我们把被误分的样本点也选择出来

PSVM样本选择技术

假设我们已经对数据集 $A_c \in R^{m_c \times m}$ 及其类别信息 $D_c \in R^{m_c \times m_c}$ 进行训练得到了分类器：$f_c(x)=\mathrm{sgn}\left(K(x',A_c')(Du)_c-r_c\right)$，现在到来了一批新的数据 $A_n \in R^{m_n \times m}$ 及其类别信息 $D_n \in R^{m_n \times m_n}$ 需要加入到分类器中。我们用 $A_r \in R^{m_r \times m}$ 表示从历史数据中选择出的样本， $A_w \in R^{m_w \times m}$ 表示从新数据中选择出的样本，$A_m \in R^{m_m \times m}$ 表示错分的样本。初始时我们有如下初值：$A_c=\varnothing, A_r=\varnothing, A_m=\varnothing, (Du)_c=0, r_c=0$。最后，令 num 表示每一步需要选择出的样本的个数。

PSVM样本选择技术

1) 等待，直到有新的数据集 A_n 需要学习；

2) 如果 $A_c \neq \varnothing$ ，令 $A_n = A_n \cup A_m$ 并跳转到第 7 步，否则令 $A_c = A_n$ ；

3) 利用公式(3.16)、(3.17)求得 $(Du)_c, r_c$ 的解；

4) 按照样本点 Lagrange 对偶变量 $(Du)_c$ 的绝对值递增的顺序选择前 num 个样本构造数据集 A_r ，并令 $A_c = A_r$ ；

5) 利用数据集 A_c 重新训练分类器，并得到 $(Du)_c, r_c$ 的新的解；

6) 利用分类器 $f_c(x)$ 对全部样本进行测试，获得错分的样本 A_m ，转移到步骤 1；

7) 通过下述公式获得新数据集 A_n 距离当前拟合超平面的距离向量 d ：

PSVM样本选择技术

$$d = \left| f(A_n) \right| = \left| \mathrm{sgn}\left(K(A_n{}', A_c{}')(Du)_c - r_c \right) \right| \qquad (4.1)$$

8） 设定一个界 d_{bound} 作为样本点与拟合超平面距离的界，一个可能的选择是 $(Du)_c$ 中元素绝对值的最大值；

9） 将数据集 A_n 压缩至 A_w ，仅保留对应距离向量 d 中元素 d_i 满足 $d_i > d_{bound}$ 条件的点；

10）将 A_w 作为增量学习中新增数据集，利用增量学习算法得到 $(Du)_c, r_c$ 的新的解，即令公式(3.20)、(3.21)、(3.22)中 $A_1 = A_c, A_2 = A_w, A_c = A_3$；跳转到第4步（此时 A_w 已经包含进 A_c ）。

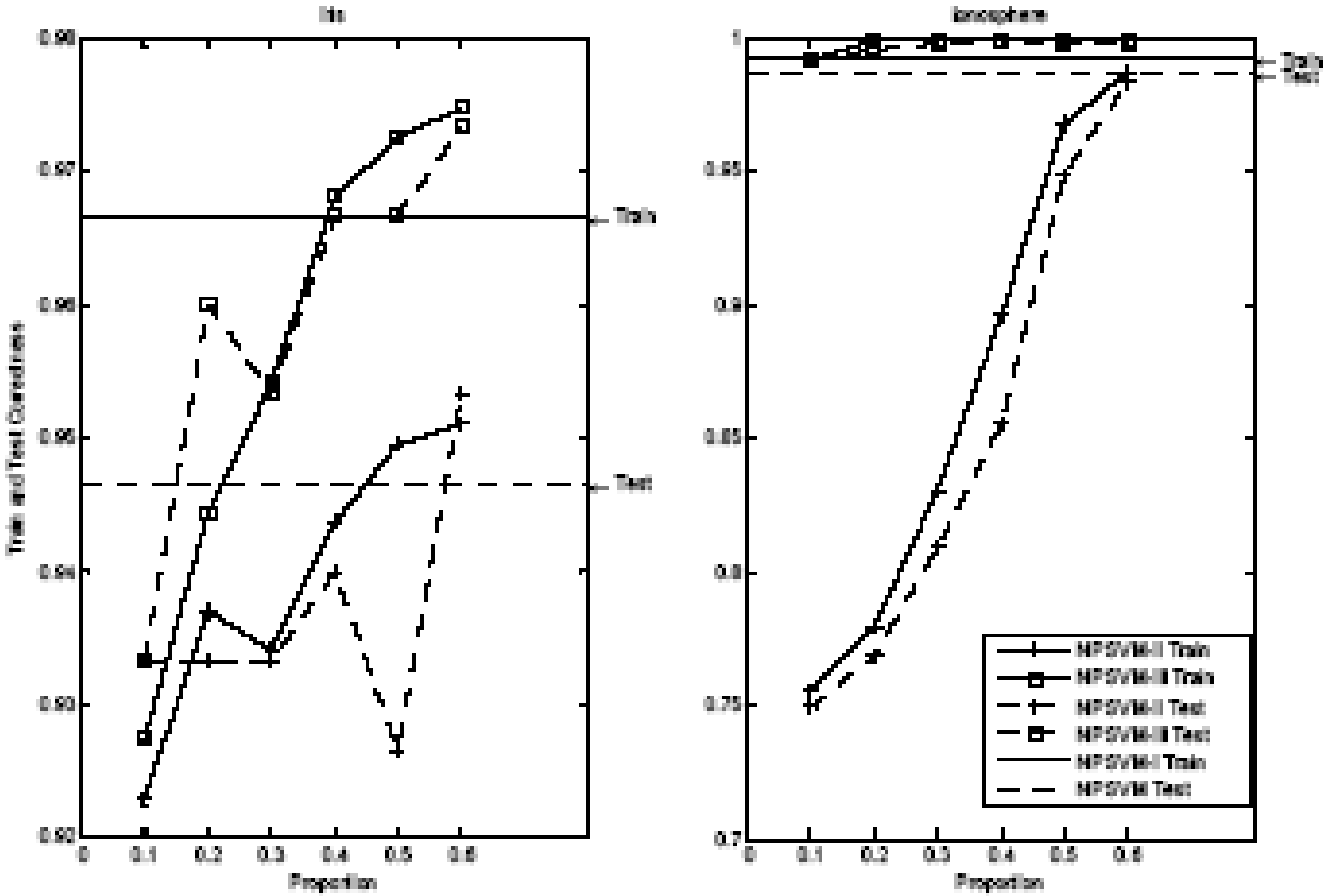
PSVM样本选择技术

Data Set $m \times n$		Iris 150×4	Ionospher e 351×34	Tic-Tac-Toe 958×9
PSVM	Train	73.93%	90.76%	98.33%
	Test	71.33%	86.08%	98.33%
PSVM-I(0.1)	Train	76.52%	84.74	98.33%
	Test	74.67%	82.05	98.32%
PSVM-I(0.2)	Train	75.70%	89.43%	98.33%
	Test	73.33%	86.06%	98.33%
PSVM-I(0.3)	Train	73.93%	89.30%	98.33%
	Test	71.33%	86.63%	98.32%

PSVM样本选择技术

Data Set $m \times n$	Iris 150×4	Ionosphere 351×34	Tic-Tac-Toe 958×9	BUPA Liver 345×6
RSWM	95.33 %	92.31%	98.50%	74.86%
NPSVM-III (10%)	95.33 %	93.44%	98.23%	68.69%
NPSVM-III (20%)	97.33 %	95.15%	98.54%	75.65%
NPSVM-III (30%)	**98.00 %**	**95.73%**	**98.64%**	**77.39%**

PSVM样本选择技术



PSVM样本选择技术

- 实验显示上述样本选择技术仅需付出较小时间代价，就可以有效地处理大样本集的在线学习问题，而且可以得到与利用全部样本进行训练的结果相近的正确率。

PSVM样本选择技术

- Qiu-ge Liu, Qing He, Zhong-zhi Shi. Data Selection for Nonlinear Proximal Support Vector Machine, Third International Conference on Natural Computation，Vol.1,pp.120-124.（EI已收录）

目录

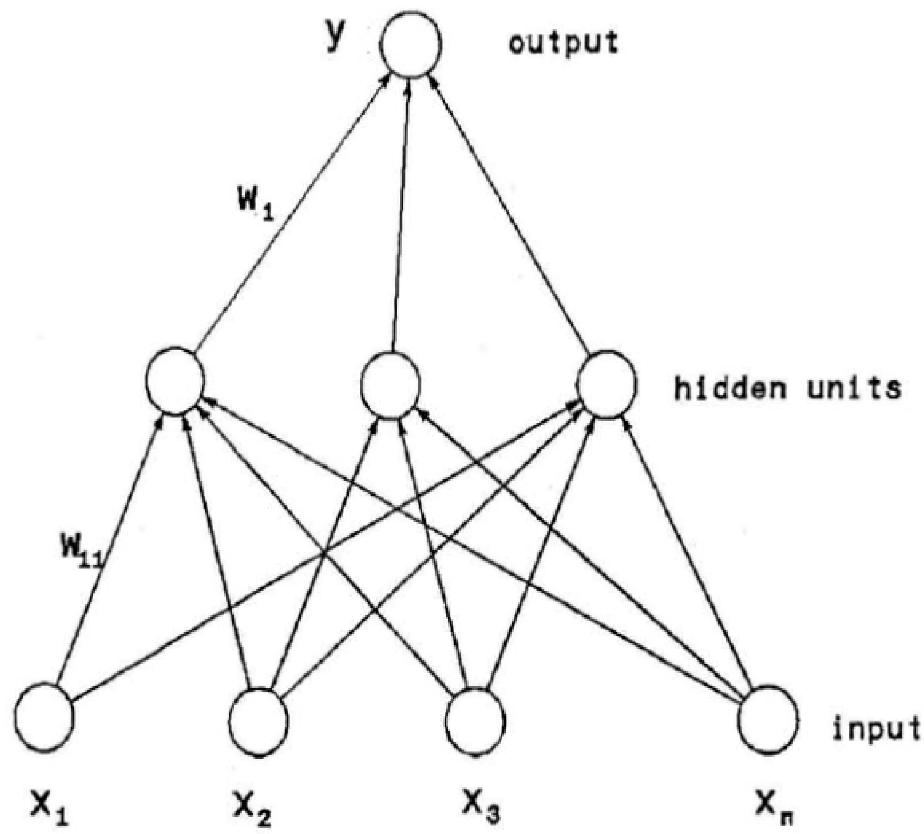
- 引言
- 非线性PSVM增量学习算法
- PSVM样本选择技术
- 极端支持向量：ESVM

极端支持向量：ESVM

- 我们提出了一种新的非线性SVM学习方法：Extreme Support Vector Machine（ESVM）
- 它基于随机构造的映射函数和正则化网络得到非线性模型
- 优点：
  - 与SVM比：运算速度快，扩展能力强，相当的正确率
  - 与SFLNSs比：更好的泛化性能

极端支持向量：ESVM

- 单隐层神经网络训练算法Extreme Learning Machine：ELM
  - 随机确定输入参数
  - 解析求解输出参数
- 速度快，避免局部极小值



极端支持向量：ESVM

- ELM两个步骤
  1. 输入数据被映射到隐层输出向量
  2. 由隐层输出向量解析求解隐层输出权重
- 同Nonlinear SVM中的核思想类似
- 神经网络的学习能力

极端支持向量：ESVM

$$\begin{aligned} \min_{(w,\gamma,y)\in R^{n+1+m}} \quad & \nu\frac{1}{2}\|y\|^2 + \frac{1}{2}(w'w+\gamma^2) \\ \text{s.t.} \quad & D(Aw - e\gamma) + y = e \end{aligned}$$

$$\begin{bmatrix} w \\ \gamma \end{bmatrix} = (\frac{I}{\nu} + E'E)^{-1}E'De,$$

$$\begin{aligned} \min_{(w,r,y)\in R^{\tilde{n}+1+m}} \quad & \nu\frac{1}{2}\|y\|^2 + \frac{1}{2}\left(w'w + r^2\right), \\ s.t. \quad & D\left(\phi(A)w - er\right) + y = e. \end{aligned}$$

$$\begin{bmatrix} w \\ r \end{bmatrix} = (\frac{I}{\nu} + E_{\Phi}'E_{\Phi})^{-1}E_{\Phi}'De$$

$$\Phi(x): R^n \to R^{\tilde{n}}$$

显示的利用一个非线性转换函数

极端支持向量：ESVM

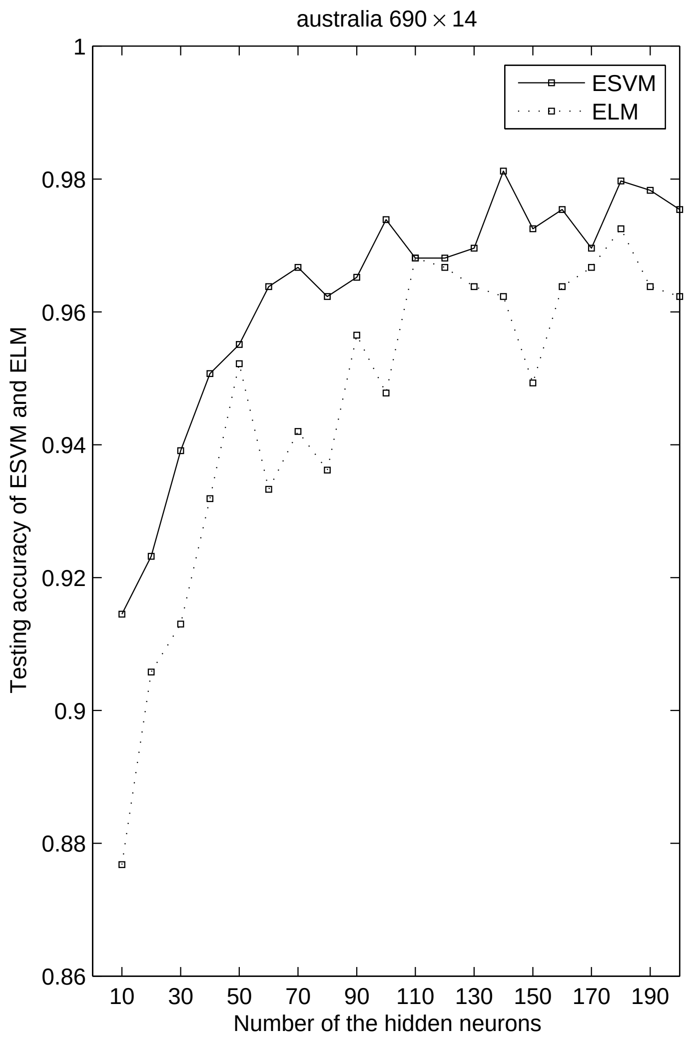
ESVM 分类器

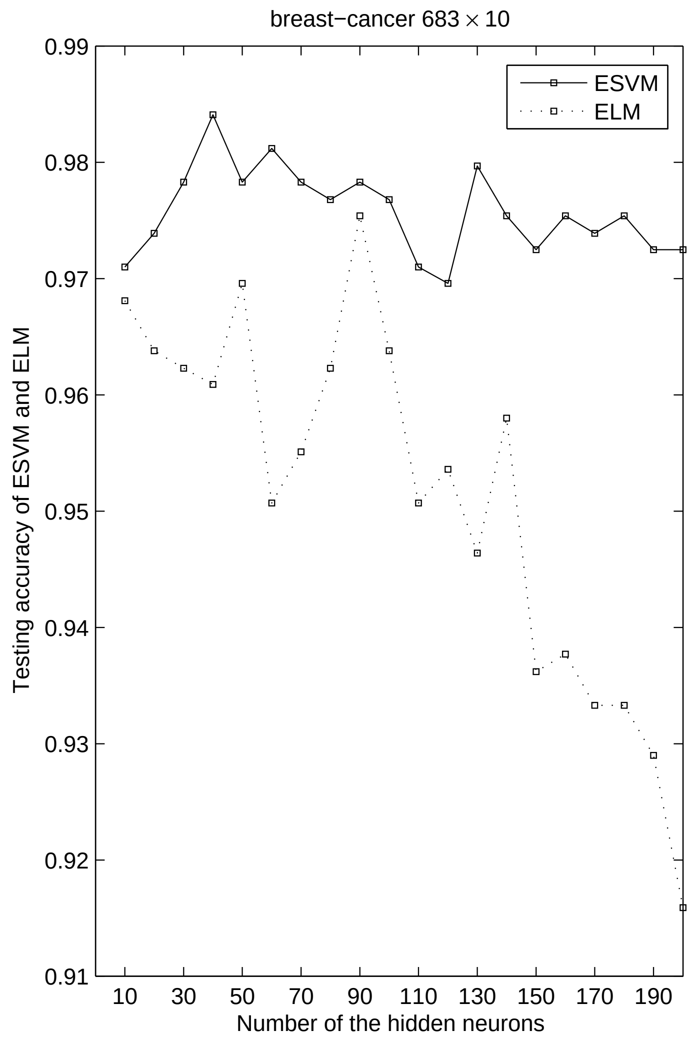
$$\Phi(x)'w-r\begin{cases} >0, \text{ then } x\in A+ \\ <0, \text{ then } x\in A- \\ =0, \text{ then } x\in A+ \text{ or } x\in A- \end{cases}$$

增量计算

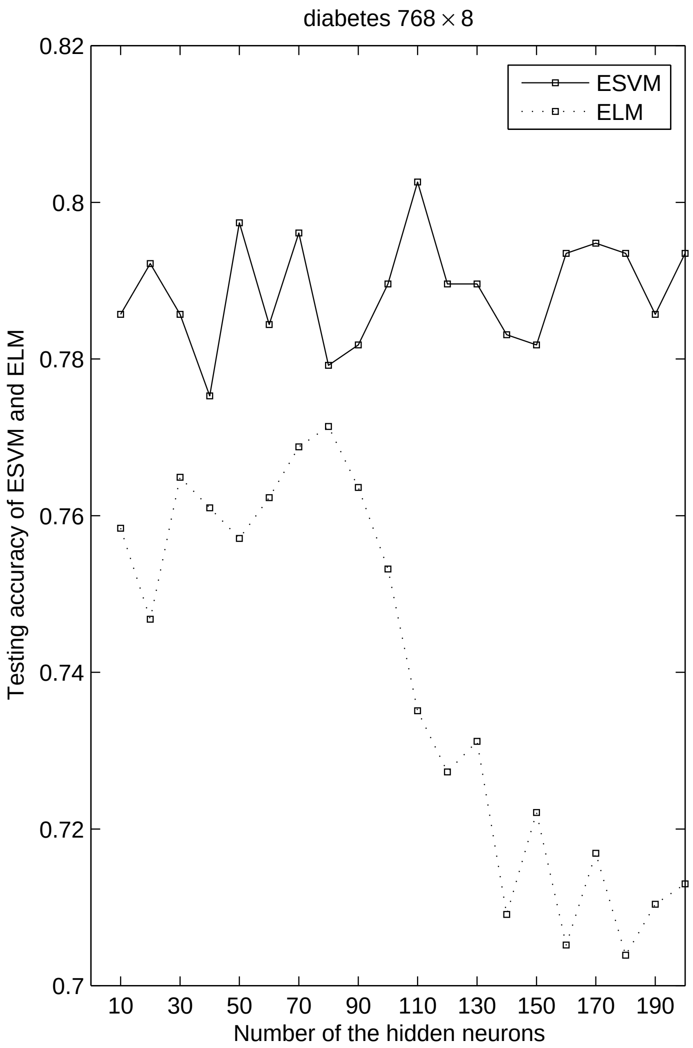
$$E_\Phi' E_\Phi = \sum E_{\Phi i}' E_{\Phi i}, E_\Phi' De = \sum E_{\Phi i}' D_i e$$

极端支持向量：ESVM

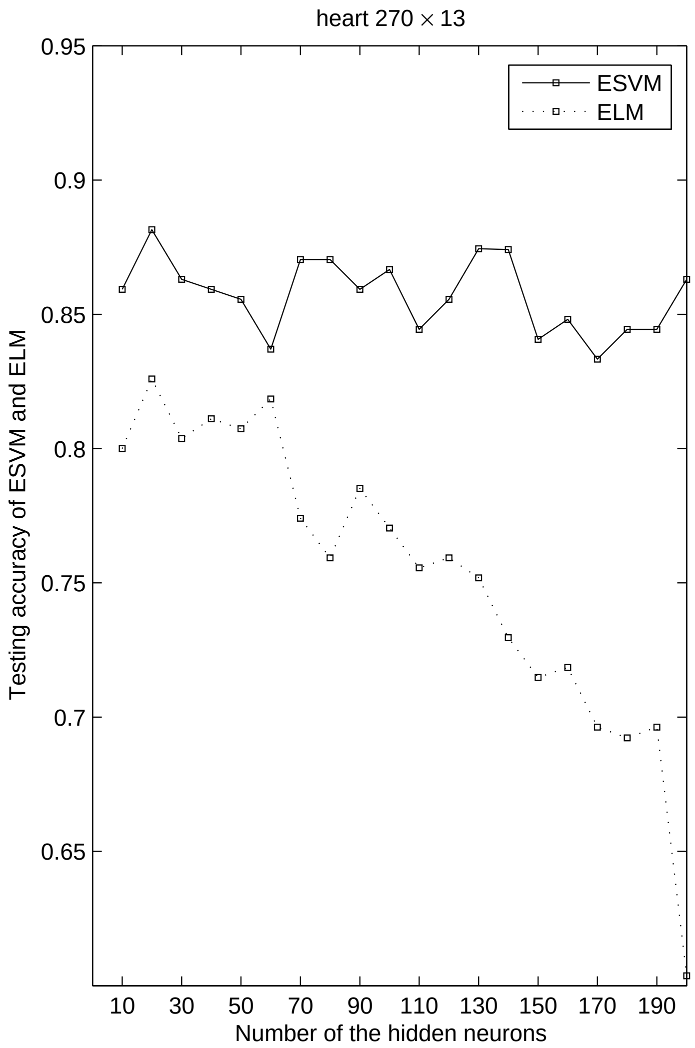

australia 690 × 14
ESVM
ELM
Testing accuracy of ESVM and ELM
Number of the hidden neurons
1
0.98
0.96
0.94
0.92
0.9
0.88
0.86
10
30
50
70
90
110
130
150
170
190


breast−cancer 683 × 10
ESVM
ELM
Testing accuracy of ESVM and ELM
Number of the hidden neurons
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0.92
0.91
10
30
50
70
90
110
130
150
170
190

极端支持向量：ESVM

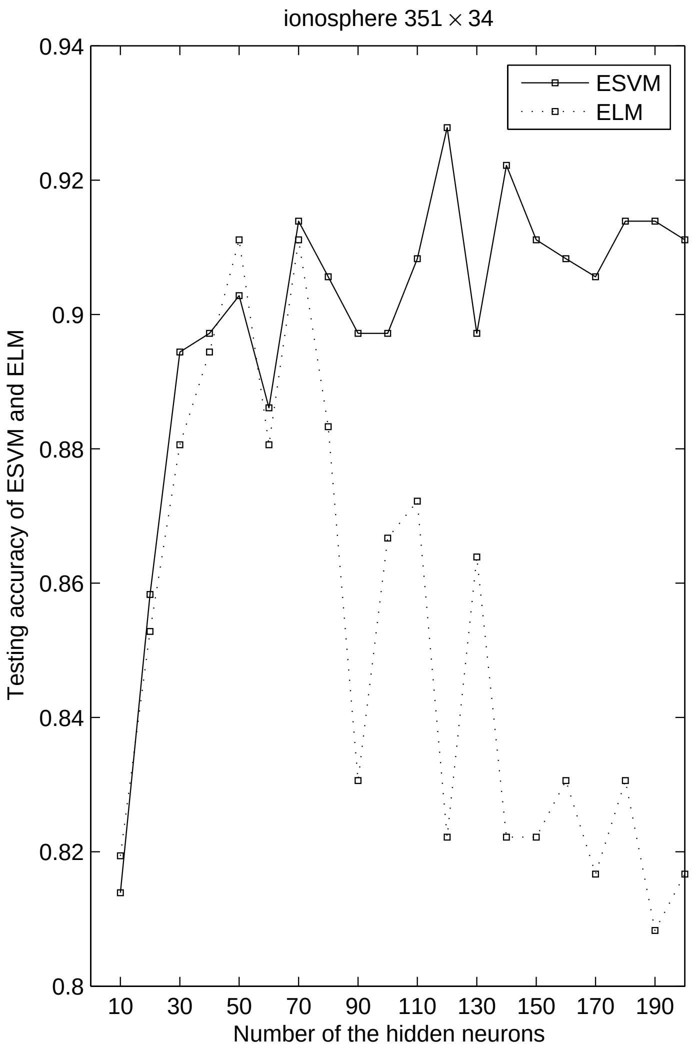


diabetes 768 × 8
ESVM
ELM
Testing accuracy of ESVM and ELM
Number of the hidden neurons

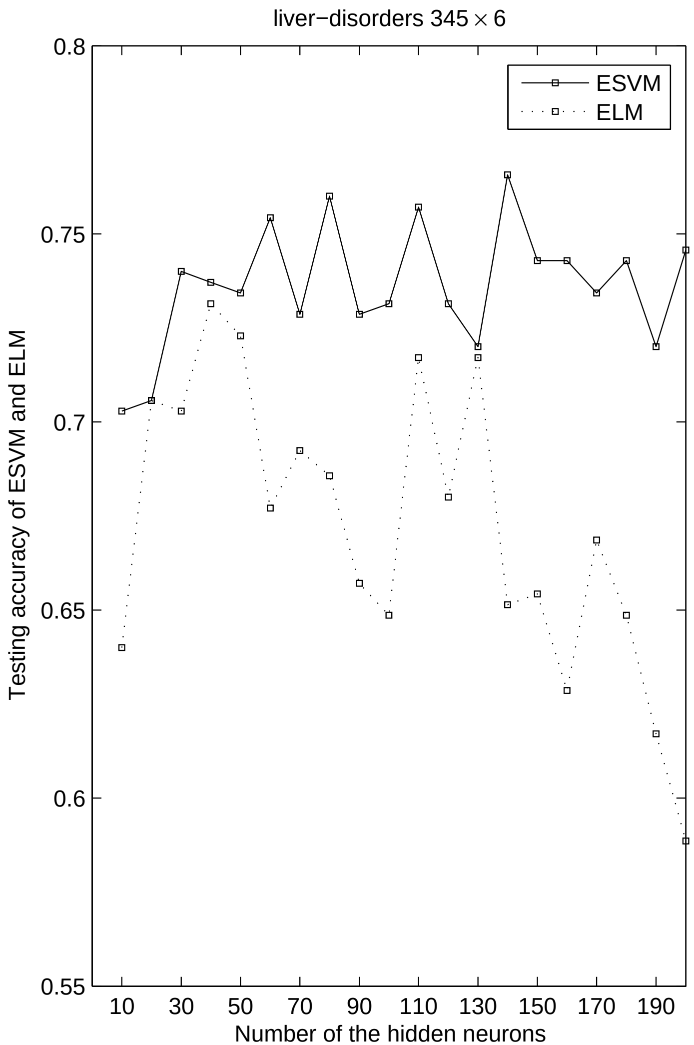


heart 270 × 13
ESVM
ELM
Testing accuracy of ESVM and ELM
Number of the hidden neurons

极端支持向量：ESVM

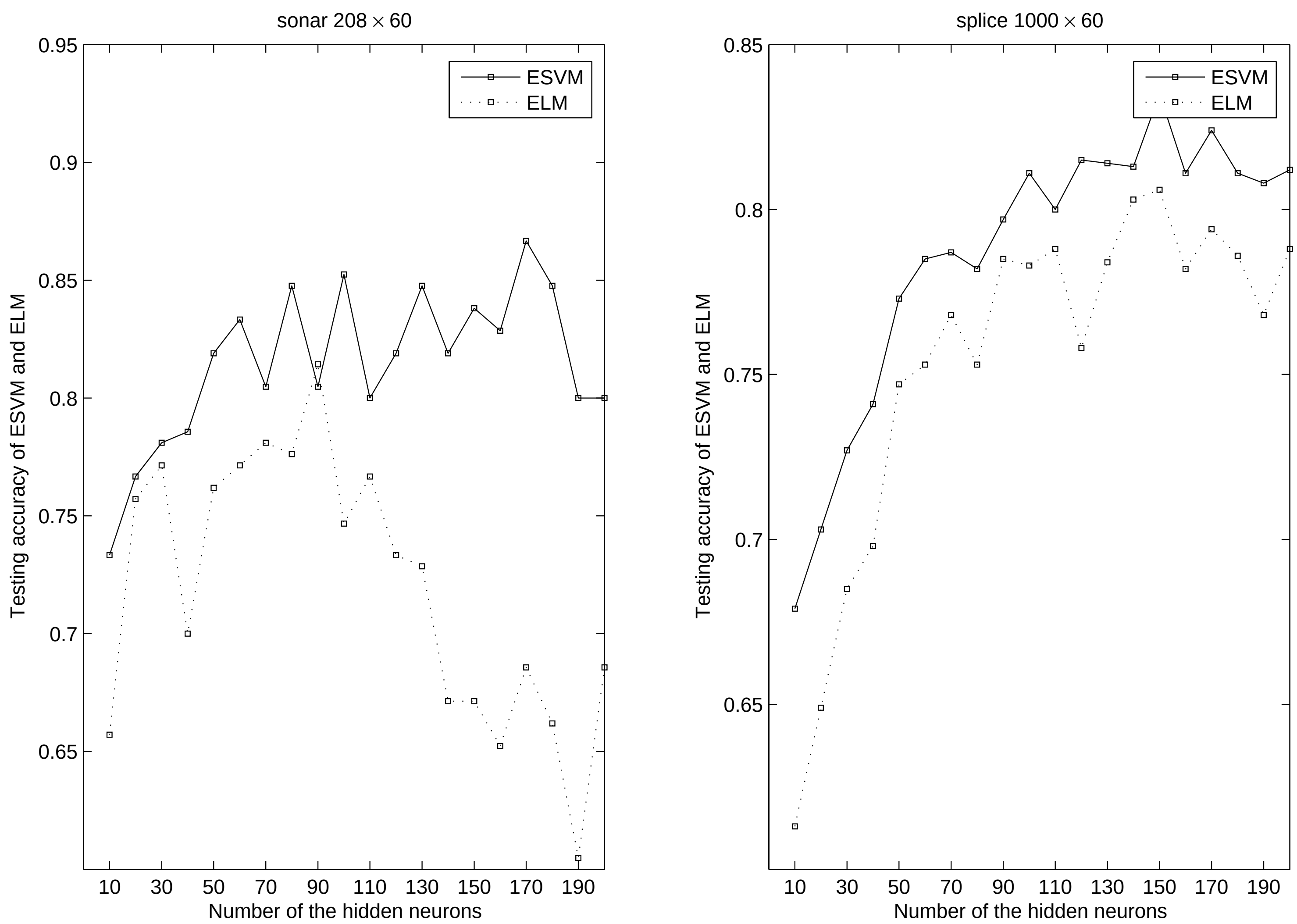


ionosphere 351 × 34
ESVM
ELM
Testing accuracy of ESVM and ELM
Number of the hidden neurons
0.94
0.92
0.9
0.88
0.86
0.84
0.82
0.8
10
30
50
70
90
110
130
150
170
190



liver−disorders 345 × 6
ESVM
ELM
Testing accuracy of ESVM and ELM
Number of the hidden neurons
0.8
0.75
0.7
0.65
0.6
0.55
10
30
50
70
90
110
130
150
170
190

极端支持向量：ESVM



sonar 208 × 60
splice 1000 × 60
ESVM
ELM
Testing accuracy of ESVM and ELM
Number of the hidden neurons
0.95
0.9
0.85
0.8
0.75
0.7
0.65
10 30 50 70 90 110 130 150 170 190

极端支持向量：ESVM

Datasets	ESVM Train Test Time						SVM Train Test Time	NPSVM Train Test Time
	20	60	100	140	180	200		
Australia 690 × 14	91.26% 92.32% 0.0047	96.09% 96.38% 0.0141	97.86% 97.39% 0.0219	98.42% 98.12% 0.0469	99.19% 97.97% 0.0703	99.28% 97.54% 0.0828	92.59% 83.91% 0.1703	100% 96.52% 0.3297
breast-cancer 683 × 10	97.08% 97.39% 0	97.54% 98.12% 0.0125	97.77% 97.68% 0.0281	97.74% 97.68% 0.0453	97.92% 97.54% 0.0672	97.85% 97.25% 0.0781	96.73% 96.63% 0.125	97.48% 97.73% 0.3281
diabetes 768 × 8	78.17% 79.22% 0.0078	80.46% 78.44% 0.0172	79.15% 78.96% 0.0313	80.81% 80.81% 0.0516	79.83% 79.83% 0.0766	85.33% 85.33% 0.0906	77.47% 75.78% 0.1689	79.15% 77.48% 0.4406
heart 270 × 13	85.76% 88.15% 0.0047	85.35% 83.70% 0.063	88.72% 86.67% 0.0109	89.26% 87.41% 0.0219	90.12% 84.44% 0.0313	84.73% 86.30% 0.0344	96.75% 75.56% 0.0312	83.29% 82.96% 0.0297
ionosphere 351 × 34	85.62% 85.83% 0.0031	94.19% 88.61% 0.0094	96.67% 89.72% 0.0156	96.32% 92.22% 0.0281	94.19% 91.39% 0.0344	97.46% 91.11% 0.0437	100% 92.02% 0.0610	99.37% 94.87% 0.0626
liver 345 × 6	75.13% 70.57% 0.0016	75.35% 75.43% 0.0063	77.23% 73.14% 0.0156	78.16% 76.57% 0.0234	76.32% 74.29% 0.0359	74.97% 74.57% 0.0453	80.58% 72.49% 0.05	76.75% 73.34% 0.0581
sonar 208 × 60	81.18% 76.67% 0.0016	90.43% 83.33% 0.0031	90.91% 85.24% 0.0141	99.89% 81.90% 0.0172	99.57% 84.76% 0.0313	87.49% 80% 0.0281	100% 74.04% 0.0405	100% 89.47% 0.0156
splice 1000 × 60	68.31% 70.30% 0.0063	80.08% 78.50% 0.0234	83.99% 81.10% 0.0484	86.63% 81.30% 0.0703	88.44% 81.10% 0.1	86.17% 81.20% 0.1141	100% 56.9% 1.25	- - -

极端支持向量：ESVM

- 我们提出了一种新的Extreme Support Vector Machine（ESVM）分类模型：
  - 利用随机构造的映射函数代替核函数
  - 正则化最小二乘
- 优点：
  - 比SLFN具有更好的泛化能力
  - 比SVM更简单，更快速
  - 具有与SVM相似的正确率
  - 可用于大数据集的训练

极端支持向量：**ESVM**

- Qiuge Liu, Qing He, and Zhongzhi Shi. Extreme Support Vector Machine, in PAKDD'08(EI源).

Most Infuluential Paper Award

Most Influential Paper Award

"Extreme Support Vector Machine Classifier"

co-authored by Qiuge Liu, Qing He and Zhongzhi Shi,
published at PAKDD 2008

研究工作总结

- Qiuge Liu, Qing He, and Zhongzhi Shi. Extreme Support Vector Machine, in PAKDD'08 (EI源).
- Qiuge Liu, Qing He, and Zhongzhi Shi. Incremental Nonlinear Proximal Support Vector Machine, in D. Liu et al. (Eds.): ISNN 2007, LNCS 4493, Part III, pp. 336–341, 2007(EI已收录).
- Qiu-ge Liu, Qing He, Zhong-zhi Shi. Data Selection for Nonlinear Proximal Support Vector Machine, Third International Conference on Natural Computation，Vol.1,pp.120-124.（EI已收录 ）
- 刘秋阁， 何清， 史忠植. 一种新的非线性支持向量机分类算法. CAAI-12, 北京邮电大学出版社, 2007, 190-195.

公开问题

- ESVM的进一步完善
  - 随机映射函数对分类器性能的影响，及其与核函数、核优化理论的关系；
  - ESVM对大数据集分类能力的分析；
  - ESVM中两步学习方式，可以推广到很多不同的（几乎任意）学习算法中用于学习非线性模型；

谢谢!