附录1：定理1和定理2的证明

定理1的整体证明思路参照Schölkopf等（2001），与之不同的是，需要重新计算带$\|\boldsymbol{\omega}\|_{2,1}\leqslant 1$约束的线性分类器类的覆盖数上界。

定义A1 设$L(\mathcal{X})$是$\mathcal{X}$上有可数支撑$supp(f)$的实值非负函数f的集合，即$L(\mathcal{X})$中的函数取值非零的点最多有可数个。令$L(\mathcal{X})$上的$\|f\|_1=\sum_{\boldsymbol{x}\in supp(f)} f(\boldsymbol{x})$，记$L^B(\mathcal{X})=\left\{f\in L(\mathcal{X}):\|f\|_1\leqslant B\right\}$。$\delta_{\boldsymbol{x}}\in L(\mathcal{X})$的定义如下：

$$\delta_{\boldsymbol{x}}(\boldsymbol{y})=\begin{cases}1, & \text{如果 } \boldsymbol{y}=\boldsymbol{x}\\ 0. & \text{其他情形}\end{cases}$$

对于函数$f\in\mathcal{F}$，训练集$\boldsymbol{X}$和$\vartheta\in\mathrm{R}$，定义函数$g_f\in L(\mathcal{X})$为：

$$g_f(\boldsymbol{y})=g_f^{\boldsymbol{X},\vartheta}(\boldsymbol{y})=\sum_{\boldsymbol{x}\in\boldsymbol{X}} d(\boldsymbol{x},f,\vartheta)\,\delta_{\boldsymbol{x}}(\boldsymbol{y})$$

函数类$\mathcal{F}$在有限样本$\boldsymbol{X}$上的ℓ_∞范数定义为：

$$\|f\|_{\ell_\infty^{\boldsymbol{x}}}=\max_{\boldsymbol{x}\in\boldsymbol{X}}|f(\boldsymbol{x})|$$

令$\mathcal{N}(\gamma,\mathcal{F},n)=\max_{\boldsymbol{X}\in\mathcal{X}^n}\mathcal{N}\left(\gamma,\mathcal{F},\ell_\infty^{\boldsymbol{X}}\right)$。

下文中，log 表示以 2 为底的对数。

定理A1（Schölkopf等，2001）假定常数$B>0$，连续型概率分布P定义在样本空间$\mathcal{X}$上，$\mathcal{F}$为sturdy实值函数类。那么，在随机抽取的大小为n的训练集$\boldsymbol{X}$上，对于使得$g_f=g_f^{\boldsymbol{X},\vartheta}\in L^B(\mathcal{X})$（即$\sum_{\boldsymbol{x}\in\boldsymbol{X}} d(\boldsymbol{x},f,\vartheta)\leqslant B$）成立的任意$\gamma>0$、$f\in\mathcal{F}$、任意$\vartheta$，以概率$1-\delta$有下列不等式成立：

$$P\{\boldsymbol{x}:f(\boldsymbol{x})<\vartheta-2\gamma\}\leqslant\frac{2}{n}\left(k+\log\frac{n}{\delta}\right)$$

其中，$k=\left\lceil\log\mathcal{N}(\gamma/2,\mathcal{F},2n)+\log\mathcal{N}(\gamma/2,L^B(\mathcal{X}),2n)\right\rceil$。

引理A1（Schölkopf等，2001）对于任意的$\gamma>0$，有

$$\log\mathcal{N}(\gamma,L^B(\mathcal{X}),n)\leqslant b\log(b^{-1}e(n+b-1))$$

其中，$b=\left\lfloor\frac{B}{2\gamma}\right\rfloor$。

引理A2 设$\mathcal{F}$为支撑为单位球的满足$\|\boldsymbol{\omega}\|_1\leqslant 1$的线性分类器类，那么

$$\log\mathcal{N}(\varepsilon,\mathcal{F},n)\leqslant c'\varepsilon^{-1}\sqrt{\log n\log p}$$

证明：基于Williamson等（2000）的引理20，有

$$e_{2k+1}(\mathcal{F},n)\leqslant ck^{-1}(\log(k^{-1}p+1)\log(k^{-1}n+1))^{\frac{1}{2}}$$

其中，$e_k(\mathcal{F},n)=\inf\{\varepsilon>0:\mathcal{N}(\varepsilon,\mathcal{F},n)\leqslant 2^{k-1}\}$。由Williamson等（2000）的引理11，求解相应的常数k可得到$\mathcal{F}$类的覆盖数上界：

$$\log\mathcal{N}(\varepsilon,\mathcal{F},n)\leqslant c'\varepsilon^{-1}\sqrt{\log n\log p}$$

引理A3 设$\mathcal{F}_{2,1}$为支撑为单位球的满足$\|\boldsymbol{\omega}\|_{2,1}\leqslant 1$的线性分类器类，那么

$$\log\mathcal{N}\left(\varepsilon,\mathcal{F}_{2,1},n\right)\leqslant c'\varepsilon^{-1}\sqrt{(k_n+q)\log n\log p}$$

证明：由于$\sqrt{k_n+q}\,\|\boldsymbol{\omega}\|_1\geqslant\|\boldsymbol{\omega}\|_{2,1}\geqslant\|\boldsymbol{\omega}\|_1$，有$\mathcal{F}_{2,1}\subset\mathcal{F}_1$，其中$\mathcal{F}_1$为支撑为单位球的满足$\|\boldsymbol{\omega}\|_1\leqslant 1$的线性分类器类。所以，

$$\begin{aligned}\mathcal{N}\left(\varepsilon,\mathcal{F}_{2,1},\|\bullet\|_{2,1}\right) &\leqslant \mathcal{N}\left(\varepsilon,\mathcal{F}_{2,1},\sqrt{k_n+q}\|\bullet\|_1\right)\\ &=\mathcal{N}\left(\varepsilon/\sqrt{k_n+q},\mathcal{F}_{2,1},\|\bullet\|_1\right)\\ &\leqslant \mathcal{N}\left(\varepsilon/\sqrt{k_n+q},\mathcal{F}_1,\|\bullet\|_1\right)\end{aligned}$$

所以，由引理A2可得，$\log\mathcal{N}\left(\varepsilon,\mathcal{F}_{2,1},n\right)\leqslant c'\varepsilon^{-1}\sqrt{(k_n+q)\log n\log p}$。

【定理1的证明】：将定理A1应用于范数不大于1的线性函数类$\mathcal{F}$。具体地，将定理A1应用于所有的

$$\log\mathcal{N}(\gamma/2,L^B(\mathcal{X}),2n)$$

其中，$B=\mathcal{D}(Z,f_\theta,0)$，以对可能的非平凡边界进行最多$n/2$次应用（每个边界使用$\delta/(n/2)$的置信度）。

记$b=\left\lfloor\frac{B}{\gamma}\right\rfloor$，显然$b=\frac{\mathcal{D}}{\gamma}$。再利用引理A1和引理A3中的覆盖数上界，则有

$$\begin{aligned}&\lceil\log\mathcal{N}(\gamma/2,\mathcal{F},2n)+\log\mathcal{N}(\gamma/2,L^B(\mathcal{X}),2n)\rceil\\ &\leqslant\left\lceil\frac{2c'}{\gamma}\sqrt{(k_n+q)\log 2n\log p}+\left\lceil b\log\left(\frac{e(2n+b-1)}{b}\right)\right\rceil\right\rceil\\ &\leqslant\frac{2c'}{\gamma}\sqrt{(k_n+q)\log 2n\log p}+b\log\left(\frac{e(2n+b-1)}{b}\right)+2\\ &=\frac{2c}{\gamma}\sqrt{(k_n+q)\log n\log p}+\frac{\mathcal{D}}{\gamma}\log\left(e\left(\frac{(2n-1)\gamma}{\mathcal{D}}+1\right)\right)+2\end{aligned}$$

假设1 协变量z_i是有界随机变量。

假设2 $\omega_j(u)\in\mathcal{H}_r$，$j=1,\cdots,p$，对于某个$r>1/2$，其中$\mathcal{H}_r$定义在$[0,1]$上，其$m$阶导数满足$\gamma\in(0,1]$阶Holder条件，$r=m+\gamma$。

假设3 协变量u定义在某紧集$[a,b]$上，其密度远离0与无穷。

【定理2的证明】：在假设1至假设3下，有

$$|f_\theta^*(z)-f_\theta(z)|\leqslant c_1 s k_n^{-r}$$

结合定理1可得定理2。

附录2：定理3的证明

由$\sum_{i=1}^{n}\alpha_i=1$，有

$$\alpha_1+\alpha_2+\cdots+\alpha_n=1 \qquad \text{（A1）}$$

其中，假设前k个$\alpha_i=\frac{1}{\nu n}$（$i=1,2,\cdots,k$），则

$$\alpha_1=\alpha_2=\cdots=\alpha_k=\frac{1}{\nu n} \qquad \text{（A2）}$$

由式（A1）、式（A2）以及$\alpha_i\in[0,\frac{1}{\nu n}]$，$i=1,2,\cdots,n$，有

$$\frac{k}{\nu n}=\alpha_1+\cdots\alpha_k=1-\alpha_{k+1}-\cdots-\alpha_n\leqslant 1$$

即有$k\leqslant\nu n$。

附录3：定理4的证明

由$\sum_{i=1}^{n}\alpha_i=1$，有

$$\alpha_1+\alpha_2+\cdots+\alpha_n=1 \tag{A3}$$

其中，假设仅有前k'个$\alpha_i\in(0,\frac{1}{\nu n}]$（$i=1,2,\cdots,k'$），则

$$\alpha_1,\alpha_2,\cdots,\alpha_{k'}\in(0,\frac{1}{\nu n}] \tag{A4}$$

由于$\alpha_i\in[0,\frac{1}{\nu n}]$，$i=1,\ 2,\cdots,n$，则剩余$n-k'$个$\alpha_i=0$，即

$$\alpha_{k'+1}=\alpha_{k'+2}=\cdots=\alpha_n=0 \tag{A5}$$

由式（A3）至式（A5），有

$$1=\alpha_1+\cdots\alpha_{k'}\leqslant\frac{k'}{\nu n}$$

即有$k'\geqslant\nu n$。

附录4：调节参数ν的选择

调节参数$\nu\in(0,1]$，为离群样本比例的上界和支持向量比例的下界。在OCSVM中，Sohn等（2021）设定ν的值为0.5。这一选择在变系数模型中是否合适及其对变量选择的准确性、分类器预测精度的影响如何需要进一步探索。

本文通过模拟实验对比不同ν值下VC-OCSVM的变量选择和预测效果。具体来说，考虑训练集样本量$n=300$，维度$p=300$，解释变量间的不同相关系数结构（0.2、0.5和0.8）以及测试集中不同异常值占比（1%、5%和10%）时的模拟效果。

表A1　　VC-OCSVM在不同ν值下的模拟结果

ν		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\tau=0.2$	TPR	1.000	1.000	1.000	1.000	0.990	0.880	0.580	0.430	0.400	0.200
	FPR	0.318	0.314	0.292	0.250	0.199	0.134	0.057	0.006	0.000	0.000
	MCC	0.186	0.187	0.198	0.219	0.250	0.274	0.287	0.526	0.623	0.444
	ROC-AUC: 1%	0.845	0.838	0.850	0.867	0.863	0.845	0.831	0.830	0.819	0.694
	ROC-AUC: 5%	0.867	0.866	0.880	0.892	0.890	0.866	0.842	0.857	0.853	0.702
	ROC-AUC: 10%	0.847	0.851	0.861	0.867	0.863	0.836	0.810	0.817	0.813	0.684
$\tau=0.5$	TPR	1.000	1.000	1.000	1.000	0.990	0.970	0.860	0.600	0.420	0.200
	FPR	0.321	0.314	0.300	0.262	0.218	0.162	0.106	0.039	0.003	0.000
	MCC	0.185	0.188	0.194	0.212	0.235	0.272	0.299	0.346	0.575	0.444
	ROC-AUC: 1%	0.816	0.801	0.829	0.832	0.830	0.817	0.807	0.790	0.764	0.648
	ROC-AUC: 5%	0.860	0.863	0.869	0.870	0.873	0.860	0.847	0.811	0.786	0.645
	ROC-AUC: 10%	0.852	0.856	0.859	0.866	0.869	0.864	0.845	0.814	0.786	0.639
$\tau=0.8$	TPR	0.980	0.980	0.980	0.980	0.980	0.960	0.930	0.890	0.750	0.090
	FPR	0.293	0.289	0.270	0.231	0.192	0.148	0.109	0.073	0.039	0.002
	MCC	0.191	0.193	0.202	0.225	0.251	0.284	0.324	0.375	0.416	0.196
	ROC-AUC: 1%	0.803	0.806	0.840	0.840	0.837	0.840	0.842	0.816	0.742	0.640
	ROC-AUC: 5%	0.838	0.814	0.855	0.862	0.865	0.863	0.852	0.829	0.749	0.566
	ROC-AUC: 10%	0.845	0.846	0.857	0.866	0.867	0.862	0.844	0.821	0.738	0.546

重复20次的模拟结果如表A1。在变量选择方面，随着ν值不断增大，VC-OCSVM的TPR和FPR不断减小，说明被选出的正确变量和冗余变量也越来越少，二者在$\nu = 0.5$时取得一个相对平衡。同时，在预测效果上，VC-OCSVM在$\nu = 0.5$时几乎表现出最佳的预测精度。因此，本文认为，调节参数ν取0.5是合理的。

附录5：模拟结果

表A2　　不同方法的模拟结果（$n = 300$，$p = 100$）

τ	模型	TPR	FPR	MCC	ROC-AUC		
					1%	5%	10%
0.2	VC-OCSVM	0.998 (0.020)	0.288 (0.046)	0.334 (0.037)	0.907 (0.087)	0.925 (0.037)	0.926 (0.027)
	L1-OCSVM	0.525 (0.117)	0.233 (0.050)	0.156 (0.058)	0.880 (0.093)	0.877 (0.048)	0.882 (0.033)
	OCSVM	— —	— —	— —	0.760 (0.136)	0.764 (0.062)	0.765 (0.048)
	Gelnet	0.470 (0.097)	0.139 (0.056)	0.218 (0.071)	0.887 (0.091)	0.889 (0.051)	0.890 (0.033)
0.5	VC-OCSVM	1.000 (0.000)	0.305 (0.045)	0.322 (0.032)	0.930 (0.072)	0.922 (0.031)	0.925 (0.025)
	L1-OCSVM	0.588 (0.104)	0.299 (0.045)	0.143 (0.054)	0.792 (0.139)	0.818 (0.051)	0.817 (0.038)
	OCSVM	— —	— —	— —	0.692 (0.168)	0.718 (0.065)	0.710 (0.049)
	Gelnet	0.484 (0.087)	0.144 (0.047)	0.217 (0.068)	0.823 (0.131)	0.840 (0.046)	0.840 (0.038)
0.8	VC-OCSVM	1.000 (0.000)	0.294 (0.045)	0.330 (0.033)	0.909 (0.093)	0.915 (0.039)	0.913 (0.025)
	L1-OCSVM	0.671 (0.117)	0.393 (0.051)	0.130 (0.056)	0.715 (0.170)	0.696 (0.078)	0.713 (0.059)
	OCSVM	— —	— —	— —	0.637 (0.167)	0.634 (0.080)	0.649 (0.062)
	Gelnet	0.515 (0.107)	0.158 (0.069)	0.220 (0.061)	0.732 (0.162)	0.716 (0.073)	0.732 (0.058)

表A3　　不同方法的模拟结果（$n=500$，$p=100$）

τ	模型	TPR	FPR	MCC	ROC-AUC 1%	ROC-AUC 5%	ROC-AUC 10%
0.2	VC-OCSVM	1.000 (0.000)	0.228 (0.045)	0.385 (0.044)	0.955 (0.041)	0.951 (0.018)	0.948 (0.015)
	L1-OCSVM	0.558 (0.110)	0.220 (0.054)	0.185 (0.065)	0.898 (0.076)	0.892 (0.035)	0.891 (0.024)
	OCSVM	— —	— —	— —	0.807 (0.108)	0.797 (0.047)	0.801 (0.033)
	Gelnet	0.525 (0.102)	0.161 (0.055)	0.223 (0.062)	0.903 (0.068)	0.900 (0.028)	0.897 (0.022)
0.5	VC-OCSVM	1.000 (0.000)	0.248 (0.048)	0.367 (0.041)	0.950 (0.043)	0.949 (0.021)	0.949 (0.016)
	L1-OCSVM	0.626 (0.119)	0.337 (0.057)	0.139 (0.058)	0.848 (0.077)	0.843 (0.043)	0.841 (0.030)
	OCSVM	— —	— —	— —	0.765 (0.096)	0.755 (0.053)	0.761 (0.031)
	Gelnet	0.505 (0.104)	0.155 (0.047)	0.216 (0.069)	0.863 (0.078)	0.863 (0.039)	0.862 (0.029)
0.8	VC-OCSVM	1.000 (0.000)	0.244 (0.048)	0.370 (0.040)	0.936 (0.046)	0.934 (0.028)	0.935 (0.017)
	L1-OCSVM	0.746 (0.121)	0.500 (0.058)	0.113 (0.057)	0.706 (0.125)	0.737 (0.048)	0.740 (0.043)
	OCSVM	— —	— —	— —	0.658 (0.133)	0.697 (0.049)	0.693 (0.045)
	Gelnet	0.549 (0.102)	0.162 (0.043)	0.233 (0.068)	0.746 (0.113)	0.773 (0.041)	0.779 (0.039)

表A4　　不同方法的模拟结果（$n=500$，$p=300$）

τ	模型	TPR	FPR	MCC	ROC-AUC 1%	ROC-AUC 5%	ROC-AUC 10%
0.2	VC-OCSVM	0.994 (0.060)	0.174 (0.024)	0.271 (0.028)	0.930 (0.051)	0.926 (0.026)	0.924 (0.023)
	L1-OCSVM	0.505 (0.087)	0.138 (0.026)	0.142 (0.036)	0.876 (0.075)	0.876 (0.035)	0.878 (0.028)
	OCSVM	— —	— —	— —	0.725 (0.103)	0.714 (0.055)	0.701 (0.037)
	Gelnet	0.445 (0.078)	0.076 (0.034)	0.193 (0.064)	0.892 (0.062)	0.893 (0.028)	0.894 (0.026)
0.5	VC-OCSVM	1.000 (0.000)	0.186 (0.024)	0.262 (0.020)	0.924 (0.065)	0.921 (0.026)	0.923 (0.021)
	L1-OCSVM	0.535 (0.100)	0.188 (0.025)	0.118 (0.035)	0.834 (0.097)	0.822 (0.041)	0.822 (0.033)
	OCSVM	— —	— —	— —	0.682 (0.115)	0.662 (0.061)	0.660 (0.038)
	Gelnet	0.450 (0.083)	0.076 (0.028)	0.191 (0.058)	0.861 (0.089)	0.855 (0.036)	0.855 (0.028)
0.8	VC-OCSVM	1.000 (0.000)	0.171 (0.021)	0.275 (0.019)	0.924 (0.054)	0.912 (0.030)	0.911 (0.019)
	L1-OCSVM	0.569 (0.098)	0.240 (0.022)	0.103 (0.031)	0.713 (0.118)	0.712 (0.054)	0.708 (0.044)
	OCSVM	— —	— —	— —	0.597 (0.124)	0.597 (0.054)	0.602 (0.040)
	Gelnet	0.481 (0.103)	0.083 (0.021)	0.189 (0.051)	0.752 (0.118)	0.750 (0.050)	0.745 (0.044)

表A5　VC-OCSVM变系数与常系数选择结果

τ	$n=300$　$p=100$		$n=500$　$p=100$		$n=500$　$p=300$	
	VC	CC	VC	CC	VC	CC
0.2	0.997 (0.033)	0.985 (0.086)	1.000 (0.000)	0.990 (0.070)	1.000 (0.000)	0.995 (0.050)
0.5	1.000 (0.000)	0.915 (0.189)	1.000 (0.000)	0.900 (0.201)	1.000 (0.000)	0.940 (0.163)
0.8	0.993 (0.047)	0.735 (0.288)	1.000 (0.000)	0.695 (0.265)	1.000 (0.000)	0.715 (0.249)

参考文献

[1] Schölkopf B, Platt J, Shawe-Taylor J, et al. Estimating the Support of High-Dimensional Distribution[J]. Neural Computation, 2001, 13(7): 1443–1471.

[2] Sohn K, Li C, Yoon J, et al. Learning and Evaluating Representations for Deep One-Class Classification[A]. International Conference on Learning Representation (ICLR)[C], 2021.

[3] Williamson R C, Smola A J, Schölkopf B. Entropy Numbers of Linear Function Classes[A]. Annual Conference Computational Learning Theory[C], 2000: 309–319.