

附录 1

附录 1 给出单个数据集上单类 SVM 可写成等式约束的具体证明。单个数据集（以第 t 个数据集为例）上的单类 SVM 的优化目标如下：

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{w}_t\|^2 - \rho_t + \frac{1}{\nu n_t} \sum_{i=1}^{n_t} \xi_{ti} \\ \text{s.t.} \quad & \xi_{ti} \geq \rho_t - \mathbf{w}_t' \mathbf{x}_{ti}, i=1, \dots, n_t \\ & \xi_{ti} \geq 0 \end{aligned} \quad (1)$$

式（1）等价于如下等式约束条件的形式：

$$\begin{aligned} \min \quad & \lambda \left(\frac{1}{2} \|\mathbf{w}_t\|^2 - \rho_t \right) + \frac{1}{n_t} \sum_{i=1}^{n_t} (a_{ti})_+ \\ \text{s.t.} \quad & a_{ti} = \rho_t - \mathbf{w}_t' \mathbf{x}_{ti}, i=1, \dots, n_t \end{aligned} \quad (2)$$

具体证明过程如下，先证（1） $\Rightarrow$ （2）：

若 $\rho_t - \mathbf{w}_t' \mathbf{x}_{ti} < 0$ ，则结合式（1）可知 $\xi_{ti} \geq 0$ ；若 $\rho_t - \mathbf{w}_t' \mathbf{x}_{ti} > 0$ ，则 $\xi_{ti} \geq \rho_t - \mathbf{w}_t' \mathbf{x}_{ti}$ ；从而得出 $\xi_{ti} \geq [\rho_t - \mathbf{w}_t' \mathbf{x}_{ti}]_+$ 。则 $\min \frac{1}{2} \|\mathbf{w}_t\|^2 - \rho_t + \frac{1}{\nu n_t} \sum_{i=1}^{n_t} \xi_{ti} \geq \min \frac{1}{2} \|\mathbf{w}_t\|^2 - \rho_t + \frac{1}{\nu n_t} \sum_{i=1}^{n_t} [\rho_t - \mathbf{w}_t' \mathbf{x}_{ti}]_+$ ，当且仅当 $\xi_{ti} = \rho_t - \mathbf{w}_t' \mathbf{x}_{ti}$ 时，“ $=$ ”成立。最小化原始目标函数式（1）等价于： $\min \frac{1}{2} \|\mathbf{w}_t\|^2 - \rho_t + \frac{1}{\nu n_t} \sum_{i=1}^{n_t} [\rho_t - \mathbf{w}_t' \mathbf{x}_{ti}]_+$ ，令 $a_{ti} = \rho_t - \mathbf{w}_t' \mathbf{x}_{ti}$ ，取 $\lambda = \nu$ ，则式（1）可推出式（2）。

再证（2） $\Rightarrow$ （1）：令 $\rho_t - \mathbf{w}_t' \mathbf{x}_{ti} = \xi_{ti}$ 且 $\xi_{ti} \geq 0$ ， $\mathbf{w}_t$ ， ρ_t ， ξ_{ti} 满足式（1）的约束条件。式（2）可写成 $\min \frac{1}{2} \|\mathbf{w}_t\|^2 - \rho_t + \frac{1}{\nu n_t} \sum_{i=1}^{n_t} [\rho_t - \mathbf{w}_t' \mathbf{x}_{ti}]_+$ ，把 $\rho_t - \mathbf{w}_t' \mathbf{x}_{ti} = \xi_{ti}$ 代入上式，由于 $\xi_{ti} \geq 0$ ，故 $[\xi_{ti}]_+ = \xi_{ti}$ ，得上式等价于 $\min \frac{1}{2} \|\mathbf{w}_t\|^2 - \rho_t + \frac{1}{\nu n_t} \sum_{i=1}^{n_t} \xi_{ti}$ ，故式（2）可推出式（1）。综上，式（1）和式（2）等价。

附录 2

对于调整参数 γ 的选择，在模拟 1 的设定下，分别选取 $\gamma=1$ 、 $\gamma=3$ 、 $\gamma=6$ 和 $\gamma=10$ ，结果如附录 2 表 1，综合考虑模型预测效果和模型聚类效果，选取 $\gamma=3$ 时模型聚类效果 RI 最优，且模型的召回率最高。不同 γ 取值下模型的准确率和召回率差别不大。

表 1 模拟 1 不同 γ 的整体结果

方法	γ	Accuracy	Precision	Recall	F ₁ Score	RI
MD-OCSVM (MCP)	$\gamma=1$	0.984	0.889	0.939	0.913	0.830
		(0.002)	(0.026)	(0.014)	(0.009)	(0.275)
MD-OCSVM (MCP)	$\gamma=3$	0.966	0.745	0.979	0.844	0.999
		(0.011)	(0.070)	(0.010)	(0.041)	(0.006)
MD-OCSVM (MCP)	$\gamma=6$	0.983	0.882	0.946	0.912	0.941
		(0.003)	(0.029)	(0.012)	(0.012)	(0.215)
MD-OCSVM (MCP)	$\gamma=10$	0.983	0.880	0.943	0.909	0.873
		(0.003)	(0.036)	(0.016)	(0.015)	(0.311)

注：表中为重复 100 次实验下的平均结果。

表 2 模拟 1 不同 γ 的各数据集结果

方法	γ	Mean	SD	Min	25%	50%	75%	Max
Accuracy								
MD-OCSVM (MCP)	$\gamma=1$	0.984	0.002	0.980	0.982	0.984	0.985	0.986
	$\gamma=3$	0.970	0.009	0.957	0.959	0.976	0.977	0.979
	$\gamma=6$	0.983	0.001	0.982	0.983	0.983	0.984	0.986
	$\gamma=10$	0.983	0.002	0.977	0.981	0.983	0.984	0.986
Precision								
MD-OCSVM (MCP)	$\gamma=1$	0.897	0.016	0.869	0.887	0.896	0.908	0.942
	$\gamma=3$	0.775	0.058	0.698	0.707	0.816	0.823	0.839
	$\gamma=6$	0.892	0.012	0.867	0.880	0.892	0.901	0.917
	$\gamma=10$	0.888	0.016	0.858	0.878	0.888	0.900	0.927
Recall								
MD-OCSVM (MCP)	$\gamma=1$	0.936	0.014	0.903	0.927	0.937	0.949	0.956
	$\gamma=3$	0.976	0.008	0.965	0.969	0.971	0.986	0.988
	$\gamma=6$	0.942	0.011	0.925	0.934	0.939	0.955	0.958
	$\gamma=10$	0.939	0.013	0.906	0.931	0.941	0.951	0.957

注：Mean 代表评价指标的均值，SD 代表标准误差，Min、Max 分别代表评价指标的最小值、最大值，25%、50%、75%代表指标对应的分位数，下同。

附录 3

模拟 4 考察的是数据集大样本量情形下的效果，由附录 3 表 3、表 4 可知，本文提出的方法在大样本量情形下的各评价指标均优于原有的 1-OCSVM 和 T-OCSVM。

表 3		模拟 4 整体结果			
方法	Accuracy	Precision	Recall	F ₁ Score	RI
1-OCSVM	0.691 (0.043)	0.220 (0.024)	0.921 (0.036)	0.354 (0.031)	-
T-OCSVM	0.646 (0.189)	0.237 (0.143)	0.889 (0.207)	0.355 (0.156)	-
MD-OCSVM (MCP)	0.748 (0.011)	0.265 (0.007)	0.999 (0.000)	0.419 (0.009)	0.504 (0.013)

注：表中为重复 20 次实验下的平均结果。

表 4		模拟 4 各数据集结果					
方法	Mean	SD	Min	25%	50%	75%	Max
Accuracy							
1-OCSVM	0.684	0.020	0.639	0.675	0.684	0.700	0.715
T-OCSVM	0.585	0.127	0.337	0.646	0.646	0.647	0.647
MD-OCSVM (MCP)	0.769	0.059	0.714	0.715	0.759	0.804	0.878
Precision							
1-OCSVM	0.255	0.028	0.201	0.253	0.264	0.271	0.284
T-OCSVM	0.217	0.042	0.128	0.237	0.237	0.238	0.238
MD-OCSVM (MCP)	0.270	0.037	0.242	0.242	0.242	0.316	0.318
Recall							
1-OCSVM	0.905	0.036	0.821	0.904	0.916	0.925	0.950
T-OCSVM	0.867	0.046	0.765	0.888	0.889	0.890	0.891
MD-OCSVM (MCP)	0.896	0.214	0.455	1.000	1.000	1.000	1.000

附录 4

附录 4 中表 5、表 6、表 7 分别展示模拟 1、模拟 2、模拟 3 在单个数据集的模型效果。

表 5 模拟 1 各数据集结果

方法	Mean	SD	Min	25%	50%	75%	Max
Accuracy							
1-OCSVM	0.700	0.019	0.664	0.686	0.699	0.715	0.739
T-OCSVM	0.806	0.001	0.804	0.805	0.806	0.807	0.809
DeepSVDD	0.842	0.016	0.806	0.829	0.843	0.856	0.868
AutoEncoder	0.887	0.008	0.869	0.880	0.887	0.893	0.904
MTL-OCSVM	0.965	0.001	0.963	0.964	0.964	0.965	0.966
MD-OCSVM (L_1)	0.982	0.001	0.980	0.981	0.982	0.982	0.986
MD-OCSVM (MCP)	0.970	0.009	0.957	0.959	0.976	0.977	0.979
MD-OCSVM (SCAD)	0.970	0.009	0.958	0.959	0.976	0.977	0.979
Precision							
1-OCSVM	0.244	0.015	0.219	0.234	0.243	0.258	0.275
T-OCSVM	0.348	0.002	0.344	0.346	0.347	0.349	0.354
DeepSVDD	0.360	0.024	0.311	0.340	0.363	0.377	0.407
AutoEncoder	0.448	0.020	0.412	0.434	0.445	0.460	0.491
MTL-OCSVM	0.726	0.004	0.717	0.723	0.725	0.728	0.735
MD-OCSVM (L_1)	0.865	0.016	0.844	0.853	0.862	0.871	0.914
MD-OCSVM (MCP)	0.775	0.058	0.698	0.707	0.816	0.823	0.839
MD-OCSVM (SCAD)	0.776	0.058	0.700	0.709	0.816	0.823	0.840
Recall							
1-OCSVM	0.911	0.015	0.865	0.903	0.911	0.921	0.939
T-OCSVM	0.914	0.002	0.910	0.912	0.914	0.915	0.921
DeepSVDD	0.887	0.028	0.812	0.866	0.887	0.909	0.937
AutoEncoder	0.868	0.019	0.820	0.856	0.870	0.882	0.899
MTL-OCSVM	0.987	0.001	0.985	0.986	0.987	0.987	0.989
MD-OCSVM (L_1)	0.960	0.007	0.939	0.957	0.961	0.966	0.969
MD-OCSVM (MCP)	0.976	0.008	0.965	0.969	0.971	0.986	0.988
MD-OCSVM (SCAD)	0.976	0.008	0.965	0.969	0.971	0.986	0.987

表 6 模拟 2 各数据集结果

方法	Mean	SD	Min	25%	50%	75%	Max
Accuracy							
1-OCSVM	0.693	0.022	0.654	0.679	0.69	0.709	0.736
T-OCSVM	0.736	0.108	0.648	0.649	0.651	0.867	0.868
DeepSVDD	0.820	0.039	0.770	0.786	0.803	0.864	0.877
AutoEncoder	0.894	0.011	0.875	0.885	0.895	0.902	0.918
MTL-OCSVM	0.919	0.037	0.886	0.889	0.890	0.963	0.965
MD-OCSVM (L_1)	0.911	0.019	0.886	0.894	0.914	0.931	0.934
MD-OCSVM (MCP)	0.915	0.011	0.895	0.905	0.919	0.925	0.933
MD-OCSVM (SCAD)	0.915	0.011	0.896	0.905	0.918	0.923	0.934
Precision							
1-OCSVM	0.234	0.024	0.203	0.215	0.221	0.260	0.275
T-OCSVM	0.330	0.116	0.232	0.235	0.237	0.470	0.472
DeepSVDD	0.312	0.081	0.214	0.243	0.266	0.404	0.431
AutoEncoder	0.469	0.026	0.428	0.452	0.467	0.490	0.534
MTL-OCSVM	0.560	0.139	0.441	0.447	0.451	0.727	0.738
MD-OCSVM (L_1)	0.531	0.053	0.458	0.480	0.536	0.583	0.601
MD-OCSVM (MCP)	0.545	0.034	0.489	0.523	0.539	0.557	0.636
MD-OCSVM (SCAD)	0.544	0.036	0.490	0.523	0.540	0.552	0.646
Recall							
1-OCSVM	0.879	0.044	0.807	0.844	0.859	0.924	0.957
T-OCSVM	0.894	0.036	0.858	0.864	0.867	0.937	0.939
DeepSVDD	0.773	0.188	0.553	0.612	0.674	0.997	0.999
AutoEncoder	0.944	0.045	0.889	0.902	0.921	0.998	1.000
MTL-OCSVM	0.930	0.034	0.896	0.902	0.905	0.971	0.974
MD-OCSVM (L_1)	0.831	0.140	0.660	0.724	0.735	1.000	1.000
MD-OCSVM (MCP)	0.883	0.098	0.741	0.813	0.824	1.000	1.000
MD-OCSVM (SCAD)	0.885	0.099	0.723	0.818	0.828	1.000	1.000

表 7

模拟 3 各数据集结果

方法	Mean	SD	Min	25%	50%	75%	Max
Accuracy							
1-OCSVM	0.687	0.044	0.559	0.676	0.698	0.714	0.743
T-OCSVM	0.551	0.273	0.021	0.436	0.531	0.724	0.954
DeepSVDD	0.823	0.035	0.737	0.807	0.822	0.852	0.878
AutoEncoder	0.881	0.030	0.812	0.857	0.893	0.901	0.914
MTL-OCSVM	0.444	0.375	0.000	0.140	0.341	0.798	1.000
MD-OCSVM (L_1)	0.964	0.028	0.905	0.944	0.977	0.986	0.991
MD-OCSVM (MCP)	0.953	0.038	0.853	0.939	0.970	0.981	0.989
MD-OCSVM (SCAD)	0.964	0.039	0.876	0.947	0.987	0.990	0.994
Precision							
1-OCSVM	0.231	0.046	0.095	0.216	0.247	0.261	0.281
T-OCSVM	0.317	0.311	0.007	0.099	0.179	0.565	0.935
DeepSVDD	0.315	0.097	0.084	0.306	0.341	0.379	0.429
AutoEncoder	0.406	0.130	0.086	0.329	0.470	0.486	0.532
MTL-OCSVM	0.330	0.390	0.000	0.067	0.119	0.568	1.000
MD-OCSVM (L_1)	0.873	0.021	0.829	0.866	0.876	0.885	0.912
MD-OCSVM (MCP)	0.883	0.026	0.821	0.864	0.882	0.899	0.948
MD-OCSVM (SCAD)	0.906	0.026	0.834	0.892	0.908	0.922	0.963
Recall							
1-OCSVM	0.870	0.132	0.449	0.828	0.930	0.958	0.984
T-OCSVM	0.572	0.279	0.042	0.426	0.538	0.761	0.989
DeepSVDD	0.816	0.284	0.138	0.660	0.999	1.000	1.000
AutoEncoder	0.791	0.312	0.100	0.572	1.000	1.000	1.000
MTL-OCSVM	0.741	0.326	0.000	0.642	0.889	0.996	1.000
MD-OCSVM (L_1)	0.734	0.354	0.003	0.508	0.935	1.000	1.000
MD-OCSVM (MCP)	0.746	0.348	0.055	0.524	0.996	1.000	1.000
MD-OCSVM (SCAD)	0.748	0.349	0.042	0.515	0.998	1.000	1.000

附录 5

表 8 url 参数实例	
url	参数
/search/search.do	ty=&dr=true&w=&f=&tid=&q=存款利率
/search/search.do	ty=&dr=true&w=&f=&tid=&q=“赢庆”系列代客理财 2017235 期
/search/search.do	ty=&dr=true&w=&f=&tid=&q=境外取款
/search/search.do	ty=&dr=true&w=&f=&tid=&q=风控
/brosperbank/personeb/ajaxPersoneb/ajaxCheckQuotaAction	mediumNo=6236231001111111¤cy=CNY&amount=5000.00&sessionUuid=edcbe87b-3e6c-44b5-860e-dae39dde0353&encryptDeviceId=M8v1x7haEEzTKt_VXWaU-No4v7LDOgq9MqpcDUj4qWq3uyLoic30aExYIvp1JSTXM8cS9_u_XCvJ8axgsgLdlWnRYYqtcIED7xzuoP1y1A9BUIK0rlTYvz2bj7ocdrZiLaNTrT5IJMRhbg9Sv1mhxxJVggM4b