机器学习理论研究导引
作业四

庄镇华 502022370071

2023 年 5 月 31 日

作业提交注意事项

(1) 本次作业提交截止时间为 **2023/06/03 23:59:59**, 截止时间后不再接收作业, 本次作业记零分;

(2) 作业提交方式: 使用此 LaTex 模板书写解答, 只需提交编译生成的 pdf 文件, 将 pdf 文件提交至南大网盘: https://box.nju.edu.cn/u/d/5a7e3aed5389469aaa57/

(3) pdf 文件命名方式: 学号-姓名-作业号-v 版本号, 例 MG1900000-张三-1-v1; 如果需要更改已提交的解答, 请在截止时间之前提交新版本的解答, 并将版本号加一;

(5) 未按照要求提交作业, 或 **pdf 命名方式不正确**, 将会被扣除部分作业分数.

1 [30pts] Rethinking Stability of SVR

教材 5.3.2 节证明了支持向量回归具有替换样本 β-均匀稳定性, 其中 $\beta=\frac{2r^2}{\lambda m}$. 试给出更紧的界, 即 $\beta=\frac{r^2}{\lambda m}$.

Proof.

支持向量回归的优化目标函数为，其中 $\boldsymbol{w}\in\mathbb{R}^d$，$\lambda$ 为正则化参数。

$$F_D(\boldsymbol{w})=\frac{1}{m}\sum_{i=1}^{m}\ell_\epsilon(\boldsymbol{w},(\boldsymbol{x}_i,y_i))+\lambda\|\boldsymbol{w}\|^2,$$

$$\ell_\epsilon(\boldsymbol{w},(\boldsymbol{x},y))=\begin{cases}0 & \textit{if} \quad |\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}-y|\leqslant\epsilon,\\ |\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}-y|-\epsilon & \textit{if} \quad |\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}-y|>\epsilon.\end{cases}$$

给定数据集 $D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\ldots,(\boldsymbol{x}_m,y_m)\}$，对任意 $k\in[m]$，令 $D'=D^{k,z_k'}$ 表示训练集 D 中第 k 个样本被替换为 $z_k'=(\boldsymbol{x}_k',y_k')$ 得到的数据集。令 $\boldsymbol{w}_D$ 和 $\boldsymbol{w}_{D'}$ 分别表示优化目标函数 $F_{D'}(\boldsymbol{w})$ 和 $F_D(\boldsymbol{w})$ 所得的最优解，即

$$\boldsymbol{w}_D\in\arg\min_{\boldsymbol{w}}F_D(\boldsymbol{w}) \qquad \boldsymbol{w}_{D'}\in\arg\min_{\boldsymbol{w}}F_{D'}(\boldsymbol{w}).$$

对任意样本 $(x,y)\in\mathcal{X}\times\mathcal{Y}$，分下面四种情况讨论:

1) 若 $\left|\boldsymbol{w}_D^{\mathrm{T}}\boldsymbol{x}-y\right|\leqslant\epsilon$ 且 $\left|\boldsymbol{w}_{D'}^{\mathrm{T}}\boldsymbol{x}-y\right|\leqslant\epsilon$，则有 $|\ell_\epsilon(\boldsymbol{w}_D,(\boldsymbol{x},y))-\ell_\epsilon(\boldsymbol{w}_{D'},(\boldsymbol{x},y))|$ $=0\leqslant r\|\boldsymbol{w}_{D'}-\boldsymbol{w}_D\|$;

2) 若 $\left|\boldsymbol{w}_D^{\mathrm{T}}\boldsymbol{x}-y\right|>\epsilon$ 且 $\left|\boldsymbol{w}_{D'}^{\mathrm{T}}\boldsymbol{x}-y\right|>\epsilon$，则有 $|\ell_\epsilon(\boldsymbol{w}_D,(\boldsymbol{x},y))-\ell_\epsilon(\boldsymbol{w}_{D'},(\boldsymbol{x},y))|$ $=\left||\boldsymbol{w}_D^{\mathrm{T}}\boldsymbol{x}-y|-|\boldsymbol{w}_{D'}^{\mathrm{T}}\boldsymbol{x}-y|\right|\leqslant\left|(\boldsymbol{w}_{D'}-\boldsymbol{w}_D)^{\mathrm{T}}\boldsymbol{x}\right|\leqslant r\|\boldsymbol{w}_{D'}-\boldsymbol{w}_D\|$;

3) 若 $\left|\boldsymbol{w}_D^{\mathrm{T}}\boldsymbol{x}-y\right|>\epsilon$ 且 $\left|\boldsymbol{w}_{D'}^{\mathrm{T}}\boldsymbol{x}-y\right|\leqslant\epsilon$，则有 $|\ell_\epsilon(\boldsymbol{w}_D,(\boldsymbol{x},y))-\ell_\epsilon(\boldsymbol{w}_{D'},(\boldsymbol{x},y))|=$ $\left||\boldsymbol{w}_D^{\mathrm{T}}\boldsymbol{x}-y|-\epsilon\right|\leqslant\left||\boldsymbol{w}_D^{\mathrm{T}}\boldsymbol{x}-y|-|\boldsymbol{w}_{D'}^{\mathrm{T}}\boldsymbol{x}-y|\right|\leqslant|(\boldsymbol{w}_{D'}-\boldsymbol{w}_D)^{\mathrm{T}}\boldsymbol{x}|\leqslant r\|\boldsymbol{w}_{D'}-\boldsymbol{w}_D\|$;

4) 若 $\left|\boldsymbol{w}_D^{\mathrm{T}}\boldsymbol{x}-y\right|\leqslant\epsilon$ 且 $\left|\boldsymbol{w}_{D'}^{\mathrm{T}}\boldsymbol{x}-y\right|>\epsilon$，同理可得 $|\ell_\epsilon(\boldsymbol{w}_D,(\boldsymbol{x},y))-\ell_\epsilon(\boldsymbol{w}_{D'},(\boldsymbol{x},y))|\leqslant r\|\boldsymbol{w}_{D'}-\boldsymbol{w}_D\|$

综合上述四种情况, 对任意样本有 $|\ell_\epsilon(\boldsymbol{w}_D,(\boldsymbol{x},y))-\ell_\epsilon(\boldsymbol{w}_{D'},(\boldsymbol{x},y))|\leqslant r\|\boldsymbol{w}_{D'}-\boldsymbol{w}_D\|$,
由于任意凸函数加入正则项 $\lambda\|\boldsymbol{w}\|^2$ 变成 2λ -强凸函数, 可知函数 $F_D(\boldsymbol{w})$ 和 $F_{D'}(\boldsymbol{w})$ 是 2λ -强凸函数, 进一步有

$$\begin{aligned}F_D(\boldsymbol{w}_{D'})&\geqslant F_D(\boldsymbol{w}_D)+\lambda\|\boldsymbol{w}_D-\boldsymbol{w}_{D'}\|^2,\\ F_{D'}(\boldsymbol{w}_D)&\geqslant F_{D'}(\boldsymbol{w}_{D'})+\lambda\|\boldsymbol{w}_D-\boldsymbol{w}_{D'}\|^2.\end{aligned}$$

将两式相加可得

$$\begin{aligned}\|\boldsymbol{w}_D-\boldsymbol{w}_{D'}\|^2\leqslant&(F_D(\boldsymbol{w}_{D'})-F_D(\boldsymbol{w}_D)-F_{D'}(\boldsymbol{w}_{D'})+F_{D'}(\boldsymbol{w}_D))/2\lambda\\ =&\frac{1}{2\lambda m}(\ell_\epsilon(\boldsymbol{w}_{D'},(\boldsymbol{x}_k,y_k))-\ell_\epsilon(\boldsymbol{w}_D,(\boldsymbol{x}_k,y_k))\\ &+\ell_\epsilon(\boldsymbol{w}_D,(\boldsymbol{x}_k',y_k'))-\ell_\epsilon(\boldsymbol{w}_{D'},(\boldsymbol{x}_k',y_k')))\\ \leqslant&\frac{r}{\lambda m}\|\boldsymbol{w}_D-\boldsymbol{w}_{D'}\|.\end{aligned}$$

进而

$$\|\boldsymbol{w}_D - \boldsymbol{w}_{D'}\| \leqslant r/(\lambda m)$$

$$|\ell_\epsilon(\boldsymbol{w}_D, (\boldsymbol{x}, y)) - \ell_\epsilon(\boldsymbol{w}_{D'}, (\boldsymbol{x}, y))| \leq r^2/(\lambda m)$$

由此可知支持向量回归具有替换样本 β-均匀稳定性，其中 $\beta = r^2/(\lambda m)$。 □

2 [30pts] Generalization and Stability

对任意 $k \in [m]$, 数据集 D 和样本 $\boldsymbol{z} \in \mathcal{X} \times \mathcal{Y}$, 若算法 $\mathfrak{L}$ 满足

$$\left|\hat{R}(\mathfrak{L}_D) - \textstyle\sum_{\boldsymbol{z}' \in D^{k,\boldsymbol{z}}} \frac{\ell(\mathfrak{L}_{D^{k,\boldsymbol{z}}}, \boldsymbol{z}')}{m}\right| \leqslant \beta_1,$$
$$|R(\mathfrak{L}_D) - \mathbb{E}_{\boldsymbol{z'} \sim \mathcal{D}}[\ell(\mathfrak{L}_{D^{k,\boldsymbol{z'}}}, \boldsymbol{z'})]| \leqslant \beta_2.$$

试证明: 对任意 $\epsilon > 0$ 有

$$P_{D \sim D^m}\left(\left|R(\mathfrak{L}_D) - \hat{R}(\mathfrak{L}_D)\right| \geqslant \epsilon + \beta_2\right) \leqslant 2\exp\left(\frac{-2\epsilon^2}{m(\beta_1 + 2\beta_2)^2}\right).$$

Proof.
易知 $\widehat{R}(\mathfrak{L}_D) = \frac{1}{m}\sum_{i=1}^m \ell(\mathfrak{L}_D, z_i)$，$R(\mathfrak{L}_D) = \mathbb{E}_{z\sim\mathcal{D}}[\ell(\mathfrak{L}_D, z)]$，
则 $\widehat{R}(\mathfrak{L}_{D^{k,\boldsymbol{z}}}) = \sum_{\boldsymbol{z}' \in D^{k,\boldsymbol{z}}} \frac{\ell(\mathfrak{L}_{D^{k,\boldsymbol{z}}}, \boldsymbol{z}')}{m}$，$R(\mathfrak{L}_{D^{k,\boldsymbol{z}}}) = \mathbb{E}_{z'\sim\mathcal{D}}[\ell(\mathfrak{L}_{D^{k,\boldsymbol{z}}}, z')]$，
题设即为

$$\left|\hat{R}(\mathfrak{L}_D) - \widehat{R}(\mathfrak{L}_{D^{k,\boldsymbol{z}}})\right| \leqslant \beta_1,$$
$$|\mathbb{E}_{z\sim\mathcal{D}}[\ell(\mathfrak{L}_D, z)] - \mathbb{E}_{\boldsymbol{z'}\sim\mathcal{D}}[\ell(\mathfrak{L}_{D^{k,\boldsymbol{z'}}}, \boldsymbol{z'})]| \leqslant \beta_2.$$

首先设函数

$$\Phi(D) = \Phi(z_1, z_2, \ldots, z_m) = R(\mathfrak{L}_D) - \widehat{R}(\mathfrak{L}_D)$$

对于任意 $k \in [m]$，由题设可得

$$\mathbb{E}_D[\Phi(D)] = \mathbb{E}_D[R(\mathfrak{L}_D) - \widehat{R}(\mathfrak{L}_D)] \leqslant \beta_2$$

给定样本 $\boldsymbol{z} \in X \times Y$，有

$$\left|\Phi(D) - \Phi(D^{k,z})\right| \leqslant |R(\mathfrak{L}_D) - R(\mathfrak{L}_{D^{k,z}})| + \left|\widehat{R}(\mathfrak{L}_{D^{k,z}}) - \widehat{R}(\mathfrak{L}_D)\right| \leq \beta_1 + 2\beta_2.$$

将 *MCDiarmid* 不等式应用于 $\Phi(D)$，对任意 $\epsilon > 0$ 有

$$\begin{aligned}
P_{D\sim\mathcal{D}^m}\left(\left|R(\mathfrak{L}_D) - \hat{R}(\mathfrak{L}_D)\right| \geqslant \beta_2 + \epsilon\right) &= 2P_{D\sim\mathcal{D}^m}\left(R(\mathfrak{L}_D) - \hat{R}(\mathfrak{L}_D) \geqslant \beta_2 + \epsilon\right) \\
&= 2P_{D\sim\mathcal{D}^m}(\Phi(D) \geqslant \beta_2 + \epsilon) \\
&\leqslant 2P_{D\sim\mathcal{D}^m}(\Phi(D) - \mathbb{E}[\Phi(D)] \geqslant \epsilon) \\
&\leqslant 2\exp\left(\frac{-2\epsilon^2}{m(\beta_1 + 2\beta_2)^2}\right)
\end{aligned}$$

□

3 [40pts] Consistent Surrogate Loss

考虑对率函数 $\phi(t)=\log(1+\mathrm{e}^{-t})$, 回答并证明下述问题.

1. **[10pts]** 试求解最优实值输出函数 $f_{\phi}^{*}(\boldsymbol{x})$.
2. **[15pts]** 试求解最优实值输出函数对应的最优替代泛化风险 R_{ϕ}^{*}.
3. **[15pts]** 证明对率函数针对原 0/1 目标函数具有替代一致性.

Proof.

1) 对于样本 $(\boldsymbol{x},y)\sim\mathcal{D}$, 基于样本空间和标记空间的联合分布，可得到条件概率

$$\eta(\boldsymbol{x})=P(y=+1|\boldsymbol{x})$$

给定替代函数 ϕ，它在数据分布上 $\mathcal{D}$ 上的替代泛化风险为

$$\begin{aligned}R_{\phi}(f)&=\mathbb{E}_{(x,y)\sim D}[\phi(yf(x))]\\&=\mathbb{E}_{x\sim\mathcal{D},x}[\eta(\boldsymbol{x})\phi(f(\boldsymbol{x}))+(1-\eta(\boldsymbol{x}))\phi(-f(\boldsymbol{x}))].\end{aligned}$$

进一步得到最优替代泛化风险

$$R_{\phi}^{*}=\mathbb{E}_{x\sim\mathcal{D}_{\mathcal{X}}}\left[\min_{f(x)\in\mathbb{R}}(\eta(\boldsymbol{x})\phi(f(\boldsymbol{x}))+(1-\eta(\boldsymbol{x}))\phi(-f(\boldsymbol{x})))\right]$$

从而得到替代函数的最优实值输出函数为

$$\begin{aligned}f_{\phi}^{*}(\boldsymbol{x})&=\arg\min_{f(\boldsymbol{x})\in\mathbb{R}}(\eta(\boldsymbol{x})\phi(f(\boldsymbol{x}))+(1-\eta(\boldsymbol{x}))\phi(-f(\boldsymbol{x})))\\&=\arg\min_{f(\boldsymbol{x})\in\mathbb{R}}\left(\eta(\boldsymbol{x})\log\left(1+\mathrm{e}^{-f(\boldsymbol{x})}\right)+(1-\eta(\boldsymbol{x}))\log\left(1+\mathrm{e}^{f(\boldsymbol{x})}\right)\right)\end{aligned}$$

令 $h(t)=\eta(\boldsymbol{x})\log(1+\mathrm{e}^{-t})+(1-\eta(\boldsymbol{x}))\log(1+\mathrm{e}^{t})$，则 $h'(t)=1/(1+\mathrm{e}^{-t})-\eta(\boldsymbol{x})$;
令 $h'(t)=0$, 得 $t=\log\frac{\eta(\boldsymbol{x})}{1-\eta(\boldsymbol{x})}$;
又因为 $h''(t)=\mathrm{e}^{-t}/(1+\mathrm{e}^{-t})^{2}\geq0$，所以在 $t=\log\frac{\eta(\boldsymbol{x})}{1-\eta(\boldsymbol{x})}$ 处取得最小值，可得

$$f_{\phi}^{*}(\boldsymbol{x})=\arg\min_{f(\boldsymbol{x})\in\mathbb{R}}h(f(\boldsymbol{x}))=\log\frac{\eta(\boldsymbol{x})}{1-\eta(\boldsymbol{x})}$$

2) 最优替代泛化风险为

$$\begin{aligned}R_{\phi}^{*}&=\mathbb{E}_{\boldsymbol{x}\sim\mathcal{D}_{\mathcal{X}}}\left[\min_{f(\boldsymbol{x})\in\mathbb{R}}(\eta(\boldsymbol{x})\phi(f(\boldsymbol{x}))+(1-\eta(\boldsymbol{x}))\phi(-f(\boldsymbol{x})))\right]\\&=\mathbb{E}_{\boldsymbol{x}\sim\mathcal{D}_{\mathcal{X}}}\left[\eta(\boldsymbol{x})\phi\left(\log\frac{\eta(\boldsymbol{x})}{1-\eta(\boldsymbol{x})}\right)+(1-\eta(\boldsymbol{x}))\phi\left(-\log\frac{\eta(\boldsymbol{x})}{1-\eta(\boldsymbol{x})}\right)\right]\\&=\mathbb{E}_{\boldsymbol{x}\sim\mathcal{D}_{\mathcal{X}}}[|-\eta(\boldsymbol{x})\log\eta(\boldsymbol{x})-(1-\eta(\boldsymbol{x}))\log(1-\eta(\boldsymbol{x}))|]\end{aligned}$$

3) **定理 *6.1*** 对替代函数 ϕ，若最优实值输出函数满足 $f_{\phi}^{*}\in\mathcal{F}^{*}$，且存在 $c>0$ 和 $s\geq1$ 使

$$|\eta(\boldsymbol{x})-1/2|^{s}\leqslant c^{s}(\phi(0)-\eta(\boldsymbol{x})\phi(f_{\phi}^{*}(\boldsymbol{x}))-(1-\eta(\boldsymbol{x}))\phi(-f_{\phi}^{*}(\boldsymbol{x})))$$

则替代函数 ϕ 具有一致性。

下面开始证明，易知 $\phi(0)-\eta(\boldsymbol{x})\phi(f_\phi^*(\boldsymbol{x}))-(1-\eta(\boldsymbol{x}))\phi(-f_\phi^*(\boldsymbol{x})=$
$\log 2+\eta(\boldsymbol{x})\log\eta(\boldsymbol{x})+(1-\eta(\boldsymbol{x}))\log(1-\eta(\boldsymbol{x})))$,

不妨设

$$\begin{aligned}h(\eta(\boldsymbol{x}))&=1^2\cdot(\phi(0)-\eta(\boldsymbol{x})\phi(f_\phi^*(\boldsymbol{x}))-(1-\eta(\boldsymbol{x}))\phi(-f_\phi^*(\boldsymbol{x})))-|\eta(\boldsymbol{x})-1/2|^2\\&=\log 2+\eta(\boldsymbol{x})\log\eta(\boldsymbol{x})+(1-\eta(\boldsymbol{x}))\log(1-\eta(\boldsymbol{x}))-(\eta(\boldsymbol{x})-1/2)^2\end{aligned}$$

则 $h'(\eta(\boldsymbol{x}))=\log\eta(\boldsymbol{x})-\log(1-\eta(\boldsymbol{x}))-2\eta(\boldsymbol{x})+1$,

则 $h''(\eta(\boldsymbol{x}))=\eta(\boldsymbol{x})/(1-\eta(\boldsymbol{x}))+(1-\eta(\boldsymbol{x}))/\eta(\boldsymbol{x})>0$ 恒成立,

又因为 $h'(1/2)=0$, 所以 $h(\eta(\boldsymbol{x}))\geq h(1/2)=0$,

设 $c=1$，$s=2$，基于定理 *6.1* 可知对率函数针对原 0/1 函数具有替代一致性。

□