机器学习理论研究导引
作业四

陈晟 MG21330006

2022 年 5 月 7 日

作业提交注意事项

(1) 本次作业提交截止时间为 **2022/05/17 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 文件上传到以下 ftp 服务器的指定位置：
地址：sftp://210.28.132.67:22，用户名：mlt2022，密码：mltspring2022@nju
文件夹位置：/C:/Users/mlt2022/hw_submissions/hw4_submission/ ；

(3) pdf 文件命名方式：学号-姓名-作业号-v 版本号, 例 MG1900000-张三-4-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；

(4) 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 [50pts] 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.

支持向量机是一类经典的机器学习分类方法，对二分类支持向量机而言考虑示例空间 $\chi \subseteq \mathcal{R}^d$, 标记空间 $mathcalY = -1, +1$, 以及训练集 $D = \{(x_1, y_1), (x_2, y_2)...(x_m, y_m)\}$ 给定样例 $(x, y) \in \chi \times \mathcal{Y}$ 和 $\omega \in \mathcal{R}^d$, 考虑 $hinge$ 函数, 有:

$\ell_{hinge}(\omega, (x, y)) = max(0, 1 - y\omega^T x)$

为了便于讨论，考虑未使用和函数的支持向量机，即目标函数:

$F_D(\omega) = \frac{1}{m}\sum_m^{i=1} max(0, 1 - y_i\omega^T x_i) + \lambda||\omega|^2|$

给定数据集 $D = \{(x_1, y_1), (x_2, y_2)...(x_m, y_m)\}$，对任意 $k \in [m]$, 令 $D' = D^{k, z'_k}$ 表示训练集 D 中第 k 个样例被替换为 $z'_k = (\hat{x_k}, \hat{y_k})$ 得到的数据集。令 W_D 和 $W_{\hat{D}}$ 分别表示优化目标函数 $F_D(\omega)$ 和 $F_{D'}(\omega)$ 所得的函数，即 $\omega_D \in argmin_\omega F_D(\omega)$ 和 $\omega_{D'} \in argmin_\omega F_{D'}(\omega)$

对任意样例 (x, y), 根据 *Cauchy-Schwarz* 不等式有 $|max(0, 1 - y\omega_D^T x) - max(0, 1 - y\omega_{D'}^T x| \leq |\omega_D^T x - \omega_{D'}^T x|$

$= |(\omega_D - \omega_{D'})x| \leq r||\omega_D - \omega_{D'}||$

$= \frac{1}{2\lambda m}[\ell_{hinge}(\omega_{D'}, (x_k, y_k)) - \ell_{hinge}(\omega_D, (x_k, y_k)) + \ell_{hinge}(\omega_D, (x'_k, y'_k))] \leq \frac{r}{\lambda m}||\omega_D - \omega_{D'}||$

由于任意凸函数加入正则项 $\lambda||\omega||^2$ 变成 2λ-强凸函数，由 $F_D(\omega) = \frac{1}{m}\sum_m^{i=1} max(0, 1 - y_i\omega^T x_i) + \lambda||\omega|^2|$ 可知函数 $F_D(\omega)$ 和 $F_{\hat{D}}(\omega)$ 是 2λ-强凸函数，进一步有

$F_D(\omega_{D'}) \geq F_D(\omega_D) + \lambda||\omega_D - \omega_{D'}||^2$

$F_{D'}(\omega_{D'}) \geq F_{D'}(\omega_D) + \lambda||\omega_D - \omega_{D'}||^2$

将上述两个式子相加可得:

$||\omega_D - \omega_{D'}||^2 \leq (F_D(\omega_{D'}) - F_D(\omega_D) - F_{D'}(\omega_{D'}) + F_{D'}(\omega_D)/2\lambda)$

$= \frac{1}{2\lambda m}[\ell_{hinge}(\omega_{D'}, (x_k, y_k)) - \ell_{hinge}(\omega_D, (x_k, y_k)) + \ell_{hinge}(\omega_D, (x'_k, y'_k))] \leq \frac{r}{\lambda m}||\omega_D - \omega_{D'}||$

求解上式可得: $||\omega_D - \omega_{D'}|| \leq \frac{r}{\lambda m}$

将上式带入 $|max(0, 1 - y\omega_D^T x) - max(0, 1 - y\omega_{D'}^T x| \leq |\omega_D^T x - \omega_{D'}^T x|$ 可得:

$|\ell_{hinge}(\omega_D, (x, y)) - \ell_{hinge}(\omega_{D'}, (x, y))| \leq \frac{r^2}{\lambda m}$

由此可知支持向量积具有替换样例 β-均匀稳定性，并且 $\beta = \frac{r^2}{\lambda m}$

2 [50pts] Generalization and Stability

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

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

试证明: 对任意 $\epsilon>0$ 有

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

Proof.

假设 $X_1, X_2 ..., X_n$ 为独立随机变量，并且这样的 $X_i \in [a_i, b_i]$, 则对于任意 $t > 0$, 有 *hoeffding* 不等式：

$P|S_n - E[S_n]| \geq t \leq exp(-\frac{2t^2}{\sum_n^{i=1}(b_i-a_i)^2})$

其中,$S_n = \frac{X_1+X_2+...+X_n}{n}$

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

$P_{D \sim \mathcal{D}^m}\left(\left|R\left(\mathfrak{L}_D\right)-\hat{R}\left(\mathfrak{L}_D\right)\right| \geqslant \epsilon+\beta_2\right)$

$= P_{D \sim \mathcal{D}^m}\left(\left|R\left(\mathfrak{L}_D\right)-\hat{R}\left(\mathfrak{L}_D\right)\right| - \beta_2 \geqslant \epsilon\right)$

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

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

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

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

由于 $\left|\hat{R}\left(\mathfrak{L}_D\right)-\sum_{\boldsymbol{z}' \in D^{k,\boldsymbol{z}}} \frac{\ell\left(\mathfrak{L}_{D^{k,\boldsymbol{z}}}, \boldsymbol{z}'\right)}{m}\right| \leqslant \beta_1$

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

$\xi_{D^k} \in (-2\beta_2, \beta_1)$

$\therefore$ 由 *hoeffding* 不等式有：$P_{D \sim \mathcal{D}^m}(|\hat{R}\left(\mathfrak{L}_D\right) - \mathbb{E}_{\boldsymbol{z} \sim \mathcal{D}}\left[\ell\left(\mathfrak{L}_{D^{k,\boldsymbol{z}}}, \boldsymbol{z}\right)\right]| \geqslant \epsilon) \leqslant 2 \exp \left(\frac{-2 \epsilon^2}{m\left(\beta_1+2 \beta_2\right)^2}\right)$

而由上面的推到可知,当上式成立时,即证 $P_{D \sim \mathcal{D}^m}\left(\left|R\left(\mathfrak{L}_D\right)-\hat{R}\left(\mathfrak{L}_D\right)\right| \geqslant \epsilon+\beta_2\right) \leqslant 2 \exp \left(\frac{-2 \epsilon^2}{m\left(\beta_1+2 \beta_2\right)^2}\right)$ 也成立