2023 秋季高级机器学习
习题四

2023.12.9

一. (30 points) **机器学习理论 I：岭回归的稳定性和泛化性**

本题分析岭回归算法的稳定性和泛化性。考虑示例空间 $\mathcal{X} \subseteq \mathbb{R}^d$ 和标记空间 $\mathcal{Y} \subseteq \mathbb{R}$, 以及训练集 $D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, y_m)\}$. 对任意示例 $(\boldsymbol{x}, y) \in \mathcal{X} \times \mathcal{Y}$ 和 $\boldsymbol{w} \in \mathbb{R}^d$ ，线性岭回归目标函数为:

$$F_D(\boldsymbol{w}) = \frac{1}{m} \sum_{i=1}^{m} \left(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i - y_i\right)^2 + \lambda \|\boldsymbol{w}\|^2. \tag{1}$$

考虑平方损失函数:

$$\ell_2(\boldsymbol{w}, (\boldsymbol{x}, y)) = \left(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x} - y\right)^2. \tag{2}$$

我们希望证明岭回归具有替换样本 β-均匀稳定性。可令 $D' = D^{k, \boldsymbol{z}'_k}$ 表示训练集 D 中第 k 个样例被替换为 $\boldsymbol{z}'_k = (\boldsymbol{x}'_k, y'_k)$ 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 $F_{D'}(\boldsymbol{w})$ 所得的解, 即

$$\boldsymbol{w}_D \in \arg\min_{\boldsymbol{w}} F_D(\boldsymbol{w}) \quad \text{和} \quad \boldsymbol{w}_{D'} \in \arg\min_{\boldsymbol{w}} F_{D'}(\boldsymbol{w}). \tag{3}$$

本题 $1 \sim 4$ 问将证明一个关于岭回归替换样本 β-均匀稳定性的定理：

定理 给定常数 $r > 0$, 若示例空间满足 $\mathcal{X} = \left\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| \leqslant r\right\}$, 平方函数 $\ell_2(\cdot, \cdot) \in [0, M]$, 则岭回归具有替换样例 β-均匀稳定性.

1. (4 points) 对任意样本 $(\boldsymbol{x}, y)$，写出要证明的岭回归替换样本 β-均匀稳定性的目标式，用含有 $\ell_2, \boldsymbol{w}_D, \boldsymbol{w}_{D'}, \boldsymbol{x}, y$ 的不等式来表示。
2. (5 points) 请你证明：在满足上述定理条件的示例空间和平方函数下，对任意样本 $(\boldsymbol{x}, y)$，有

$$|\ell_2(\boldsymbol{w}_D, (\boldsymbol{x}, y)) - \ell_2(\boldsymbol{w}_{D'}, (\boldsymbol{x}, y))| \leqslant 2r\sqrt{M} \|\boldsymbol{w}_D - \boldsymbol{w}_{D'}\| \tag{4}$$

3. (6 points) 一个结论可能会对证明有用。请你证明：$F_D(\boldsymbol{w})$ 满足 2λ-强凸函数的充要条件，即对任意 $\boldsymbol{w}, \boldsymbol{w'}$ 有

$$F_D(\boldsymbol{w'}) \geq F_D(\boldsymbol{w}) + \nabla F_D(\boldsymbol{w})(\boldsymbol{w'} - \boldsymbol{w}) + \lambda \|\boldsymbol{w'} - \boldsymbol{w}\|^2 \tag{5}$$

提示：令 $F_D(\boldsymbol{w}) = f(\boldsymbol{w}) + \lambda \|\boldsymbol{w}\|^2$，并利用 $f(\boldsymbol{w})$ 为凸函数的性质。

4. (10 points) 请你完成上述岭回归替换样本 β-均匀稳定性定理的证明，注意：我们需要确定 β 的值才能完成证明。
5. (5 points) 可以进一步得到岭回归的泛化误差界。请你证明，在满足上述岭回归替换样本 β-均匀稳定性定理的示例空间和平方函数的情况下，对于优化目标函数 $F_D(\boldsymbol{w})$ 所得的最优解 $\boldsymbol{w}_D$，对任意 $\delta \in (0, 1)$, 以至少 $1 - \delta$ 的概率有

$$R(\boldsymbol{w}_D) \leqslant \widehat{R}(\boldsymbol{w}_D) + \frac{4Mr^2}{\lambda m} + \left(\frac{8Mr^2}{\lambda} + M\right) \sqrt{\frac{\ln(1/\delta)}{2m}} \tag{6}$$

其中 $R(\cdot)$ 和 $\widehat{R}(\cdot)$ 分别表示泛化损失和经验损失。

解：

1. 由书中关于 $\beta-$ 样本稳定性的定义：岭回归目标函数（学习算法）关于平方损失函数 l_2 满足 $\beta-$ 稳定性，则对于任意样本 $(x,y)\in\mathcal{X}\times\mathcal{Y}$

   移除示例的 β-稳定性：$|l(\boldsymbol{w}_D,(x,y))-l(\boldsymbol{w}_{D^{-i}},(x,y))|\leq\beta$

   替换示例的 β-稳定性：$|l(\boldsymbol{w}_D,(x,y))-l(\boldsymbol{w}_{D'},(x,y))|\leq 2\beta$(三角不等式放缩)

   综上，本题需要的岭回归替换样本 β-样本稳定性的目标式为: 对于任意样本 $(x,y)\in\mathcal{X}\times\mathcal{Y}$

   $$|l(\boldsymbol{w}_D,(x,y))-l(\boldsymbol{w}_{D'},(x,y))|\leq 2\beta$$

2. $$\begin{aligned}|l_2(\boldsymbol{w}_D,(x,y))-l_2(w_{D'},(x,y))| &= \left|(\boldsymbol{w}_D^T\boldsymbol{x}-y)^2-(\boldsymbol{w}_{D'}^T\boldsymbol{x}-y)^2\right| \\ &= \left|(\boldsymbol{w}_D^T\boldsymbol{x}-y+\boldsymbol{w}_{D'}^T\boldsymbol{x}-y)\cdot(\boldsymbol{w}_D^T\boldsymbol{x}-\boldsymbol{w}_{D'}^T\boldsymbol{x})\right| \\ &\leq \left|(\boldsymbol{w}_D^T\boldsymbol{x}-y+\boldsymbol{w}_{D'}^T\boldsymbol{x}-y)\right|\cdot\left|\boldsymbol{w}_D^T\boldsymbol{x}-\boldsymbol{w}_{D'}^T\boldsymbol{x}\right| \\ &\leq \left(\left|\boldsymbol{w}_D^T\boldsymbol{x}-y\right|+\left|\boldsymbol{w}_{D'}^T\boldsymbol{x}-y\right|\right)\cdot\left|\boldsymbol{w}_D^T-\boldsymbol{w}_{D'}^T\right|\cdot|\boldsymbol{x}| \\ &\leq 2\cdot\sqrt{M}\cdot r\cdot\|\boldsymbol{w}_D-\boldsymbol{w}_{D'}\|\end{aligned}$$

3. 证明：

   $f(\boldsymbol{w})=\frac{1}{m}\sum_{i=1}^m(\boldsymbol{w}^Tx_i-y_i)^2$ 是关于 $\boldsymbol{w}$ 的仿射、平方求和的复合函数，由于仿射函数是凸函数，平方以及求和操作保凸，所以 $f(\boldsymbol{w})$ 是关于 $\boldsymbol{w}$ 的凸函数。

   $$\begin{aligned}F_D(\boldsymbol{w}')-F_D(\boldsymbol{w}) &= f(\boldsymbol{w}')-f(\boldsymbol{w})+\lambda\|\boldsymbol{w}'\|^2-\lambda\|\boldsymbol{w}\|^2 \\ &\geq \bigtriangledown f(\boldsymbol{w})^T(\boldsymbol{w}'-\boldsymbol{w})+\lambda\boldsymbol{w}'^T\boldsymbol{w}'-\lambda\boldsymbol{w}^T\boldsymbol{w} \\ &= \bigtriangledown F(\boldsymbol{w})^T-2\lambda\boldsymbol{w}^T(\boldsymbol{w}'-\boldsymbol{w})+\lambda\boldsymbol{w}'^T\boldsymbol{w}'-\lambda\boldsymbol{w}^T\boldsymbol{w} \\ &= \bigtriangledown F(\boldsymbol{w})^T+\lambda\boldsymbol{w}'^T\boldsymbol{w}'+\lambda\boldsymbol{w}^T\boldsymbol{w}-2\lambda\boldsymbol{w}^T\boldsymbol{w}' \\ &= \bigtriangledown F(\boldsymbol{w})^T+\lambda\|\boldsymbol{w}-\boldsymbol{w}'\|^2\end{aligned}$$

   以上不等于号使用了凸函数的一阶条件

   $$f(x)\text{是凸函数}\Leftrightarrow domf\text{是凸集，且}\forall x,y\in domf, f(y)\geq f(x)+\bigtriangledown f(x)^T(y-x)$$

4. 证明：

   原因 1: 为了使用 $\boldsymbol{w}_D$ 和 $\boldsymbol{w}_{D'}$ 对应数据集 D 和 D' 数据之间的替换关系

   原因 2: 为了使用岭回归目标函数性质（2λ 强凸），且岭回归问题是无约束优化，所以有 $\bigtriangledown F_D(w_D)=0,\bigtriangledown F_{D'}(w_{D'})=0$

   使用原因 2，以下分别令 $\boldsymbol{w}=\boldsymbol{w}_D,\boldsymbol{w}'=\boldsymbol{w}_{D'}$ 以及 $\boldsymbol{w}'=\boldsymbol{w}_D,\boldsymbol{w}=\boldsymbol{w}_{D'}$，带入岭回归目标函数 2λ-强凸的性质中

   $$\begin{aligned}F_D(\boldsymbol{w}_{D'})-F_D(\boldsymbol{w}_D)&\geq\bigtriangledown F_D^T(\boldsymbol{w}_D)(\boldsymbol{w}_{D'}-\boldsymbol{w}_D)+\lambda\|\boldsymbol{w}_{D'}-\boldsymbol{w}_D\|^2 \\ F_{D'}(\boldsymbol{w}_D)-F_{D'}(\boldsymbol{w}_{D'})&\geq\bigtriangledown F_{D'}^T(\boldsymbol{w}_{D'})(\boldsymbol{w}_D-\boldsymbol{w}_{D'})+\lambda\|\boldsymbol{w}_{D'}-\boldsymbol{w}_D\|^2\end{aligned}$$

   使用原因 2，两式相加，由于岭回归问题是无约束规划，所以有

   $$2\lambda\|\boldsymbol{w}_{D'}-\boldsymbol{w}_D\|^2\leq((F_D(\boldsymbol{w}_{D'})-F_{D'}(\boldsymbol{w}_{D'})+(F_{D'}(\boldsymbol{w}_D)-F_D(\boldsymbol{w}_D)))$$

使用原因 1，由于数据集 D 和 D' 数据之间的替换关系，有以下两等式

$$F_D(\boldsymbol{w}_{D'}) - F_{D'}(\boldsymbol{w}_{D'}) = \frac{1}{m}\left[(\boldsymbol{w}_{D'}^T\boldsymbol{x}_k - y_k)^2 - (\boldsymbol{w}_{D'}^T\boldsymbol{x}'_k - y'_k)^2\right]$$
$$F_{D'}(\boldsymbol{w}_D) - F_D(\boldsymbol{w}_D) = \frac{1}{m}\left[(\boldsymbol{w}_D^T\boldsymbol{x}'_k - y'_k)^2 - (\boldsymbol{w}_D^T\boldsymbol{x}_k - y_k)^2\right]$$

所以带入上式

$$\begin{aligned}
2\lambda\left\|\boldsymbol{w}_{D'} - \boldsymbol{w}_D\right\|^2 &\le (F_D(\boldsymbol{w}_{D'}) - F_{D'}(\boldsymbol{w}_{D'})) + (F_{D'}(\boldsymbol{w}_D) - F_D(\boldsymbol{w}_D)) \\
&= \frac{1}{m}\left[\left((\boldsymbol{w}_{D'}^T\boldsymbol{x}_k - y_k) + (\boldsymbol{w}_{D'}^T\boldsymbol{x}'_k - y'_k)\right)\cdot \boldsymbol{w}_{D'}^T(\boldsymbol{x}_k - \boldsymbol{x}'_k)\right] \\
&\qquad + \frac{1}{m}\left[\left((\boldsymbol{w}_D^T\boldsymbol{x}'_k - y'_k) + (\boldsymbol{w}_D^T\boldsymbol{x}_k - y_k)\right)\cdot \boldsymbol{w}_D^T(\boldsymbol{x}'_k - \boldsymbol{x}_k)\right] \\
&\le \frac{1}{m}(x_k - x'_k)\cdot(\boldsymbol{w}_{D'}^T - \boldsymbol{w}_D^T)2\cdot\sqrt{M}
\end{aligned}$$

左右同时消除 $\|\boldsymbol{w}_{D'} - \boldsymbol{w}_D\|$, 所以有

$$\begin{aligned}
\|\boldsymbol{w}_{D'} - \boldsymbol{w}_D\| &\le \frac{1}{2\lambda}\frac{1}{m}\|\boldsymbol{x}_k - \boldsymbol{x}'_k\|\cdot 2\cdot\sqrt{M} \\
&\le \frac{1}{2\lambda m}2r\cdot 2\cdot\sqrt{M} \\
&= \frac{2r\sqrt{M}}{\lambda m}
\end{aligned}$$

结合第二题结论

$$\begin{aligned}
|l_2(\boldsymbol{w}_D,(x,y)) - l_2(\boldsymbol{w}_{D'},(x,y))| &\le 2r\sqrt{M}\,\|\boldsymbol{w}_{D'} - \boldsymbol{w}_D\| \\
&\le \frac{4r^2M}{\lambda m} = 2\beta
\end{aligned}$$

综上，在 $\beta = \frac{2r^2M}{\lambda m}$ 时，岭回归目标函数（学习算法）关于 $l_2(\cdot,\cdot)$ 满足替换示例 $\beta-$ 均匀稳定性 $|l(\boldsymbol{w}_D,(x,y)) - l(\boldsymbol{w}_{D'},(x,y))| \le 2\beta$

5. 由书中定理：基于 $\beta-$ 均匀稳定性的 $\mathcal{L}$ 泛化误差界（同时适用于移除示例、替换示例 $\beta-$ 稳定性）有如下结论：

   $\forall m \ge 1, \forall\delta \in (0,1)$，以至少 $1-\delta$ 的概率有

$$R(\boldsymbol{w}_D) \le \hat{R}(w_D) + 2\beta + (4m\beta + M)\sqrt{\frac{ln(\frac{1}{\delta})}{2m}}$$

   由第四题结论：在 $\beta = \frac{2r^2M}{\lambda m}$ 时，岭回归目标函数（学习算法）关于 $l_2(\cdot,\cdot)$ 满足替换示例 $\beta-$ 均匀稳定性 $|l(\boldsymbol{w}_D,(x,y)) - l(\boldsymbol{w}_{D'},(x,y))| \le 2\beta$。所以带入如上定理，即可得

$$R(\boldsymbol{w}_D) \le \hat{R}(w_D) + \frac{4Mr^2}{\lambda m} + (\frac{8Mr^2}{\lambda} + M)\sqrt{\frac{ln(\frac{1}{\delta})}{2m}}$$

二. (30 points) **机器学习理论 II：基于覆盖数的泛化界**

设 $\mathcal{H}$ 为一假设空间, 假设的定义域为 $\mathcal{X}$, 值域为 $\mathcal{Y} \subset \mathbb{R}$. $\forall \epsilon > 0$, 可如下定义覆盖数 (Covering Number):

$$\mathcal{N}(\mathcal{H}, \epsilon) = \min \{k \in \mathbb{N} | \exists \{h_1, \cdots, h_k\} \subset \mathcal{H}, \text{ s.t. } \forall h \in \mathcal{H}, \exists i \in [k], \|h - h_i\|_\infty \leqslant \epsilon\}, \tag{7}$$

其中 $\|h - h_i\|_\infty = \max_{x \in \mathcal{X}} |h(x) - h_i(x)|$. 覆盖数可以衡量一个假设空间的复杂度: 覆盖数越大, 意味着这一假设空间越复杂. 本题利用覆盖数证明了平方损失下的一个泛化界, 该结论也可以说明覆盖数可以衡量假设空间的复杂度. 令 $\mathcal{D}$ 为 $\mathcal{X} \times \mathcal{Y}$ 上的一个分布, 且有标记样本根据这一分布采样得到. 定义 $h \in \mathcal{H}$ 的泛化误差为

$$R(h) = \mathbb{E}_{(x,y)\sim\mathcal{D}} \left[(h(x) - y)^2\right]. \tag{8}$$

训练集 $S = \{(x_1, y_1), \cdots, (x_m, y_m)\}$ 上的经验误差为

$$\hat{R}_S(h) = \frac{1}{m} \sum_{i=1}^{m} (h(x_i) - y_i)^2. \tag{9}$$

设 $\mathcal{H}$ 是有界的, 即 $\exists M > 0$, 使得 $\forall (x, y) \in \mathcal{X} \times \mathcal{Y}, h \in \mathcal{H}, |h(x) - y| \leqslant M$.

1. (9 points) 令 $L_S = R(h) - \hat{R}_S(h)$, 试证明 $\forall h_1, h_2 \in \mathcal{H}, S$,

$$|L_S(h_1) - L_S(h_2)| \leqslant 4M \|h_1 - h_2\|_\infty. \tag{10}$$

2. (9 points) 设 $\mathcal{H}$ 可以被 k 个子集 $\mathcal{B}_1, \cdots, \mathcal{B}_k$ 覆盖, 即 $\mathcal{H} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$. 试证明 $\forall \epsilon > 0$,

$$\Pr_{S\sim\mathcal{D}^m} \left[\sup_{h\in\mathcal{H}} |L_S(h)| \geqslant \epsilon\right] \leqslant \sum_{i=1}^{k} \Pr_{S\sim\mathcal{D}^m} \left[\sup_{h\in\mathcal{B}_i} |L_S(h)| \geqslant \epsilon\right]. \tag{11}$$

3. (12 points) 令 $k = \mathcal{N}\left(\mathcal{H}, \frac{\epsilon}{8M}\right)$, $\mathcal{B}_1, \cdots, \mathcal{B}_k$ 为 $\mathcal{H}$ 的覆盖, 其中 $\forall i \in [k]$, $\mathcal{B}_i$ 的圆心为 h_i, 半径为 $\frac{\epsilon}{8M}$. 试证明

$$\Pr_{S\sim\mathcal{D}^m} \left[\sup_{h\in\mathcal{B}_i} |L_S(h)| \geqslant \epsilon\right] \leqslant \Pr_{S\sim\mathcal{D}^m} \left[|L_S(h_i)| \geqslant \frac{\epsilon}{2}\right]. \tag{12}$$

并利用 Hoeffding 不等式证明

$$\Pr_{S\sim\mathcal{D}^m} \left[\sup_{h\in\mathcal{H}} \left|R(h) - \hat{R}_S(h)\right| \geqslant \epsilon\right] \leqslant \mathcal{N}\left(\mathcal{H}, \frac{\epsilon}{8M}\right) 2\mathrm{e}^{-\frac{m\epsilon^2}{2M^4}}. \tag{13}$$

解:

1. 以下设 $|S| = m$, 对 $\forall h_1, h_2 \in \mathcal{H}, \forall S \sim \mathcal{D}^m$

$$\begin{aligned} |L_S(h_1) - L_S(h_2)| &= \left|R(h_1) - \hat{R}_S(h_1) - (R(h_2) - \hat{R}_S(h_2))\right| \\ &\leq |R(h_1)) - R(h_2)| + \left|\hat{R}_S(h_1) - \hat{R}_S(h_2)\right| \end{aligned}$$

其中第一项

$$\begin{aligned} |R(h_1) - R(h_2)| &= \left|\mathbb{E}_{(x,y)\sim D} \left[(h_1(x) - y)^2 - (h_2(x) - y)^2\right]\right| \\ &= \mathbb{E}_{(x,y)\sim D} \left[|(h_1(x) - y + h_2(x) - y) \cdot (h_1(x) - h_2(x))|\right] \\ &\leq \mathbb{E}_{(x,y)\sim D} \left[|(h_1(x) - y + h_2(x) - y)| \cdot |h_1(x) - h_2(x)|\right] \\ &\leq \mathbb{E}_{(x,y)\sim D} \left[2M \|h_1 - h_2\|_\infty\right] \\ &= 2M \|h_1 - h_2\|_\infty \end{aligned}$$

其中第二项

$$\begin{aligned}\left|\hat{R}_S(h_1)-\hat{R}_S(h_2)\right| &= \left|\frac{1}{m}\sum_{i=1}^{m}\left((h_1(x_i)-y_i)^2-(h_2(x_i)-y_i)^2\right)\right| \\ &= \frac{1}{m}\sum_{i=1}^{m}\left|(h_1(x_i)+h_2(x_i)-2y_i)\cdot(h_1(x_i)-h_2(x_i))\right| \\ &\leq \frac{1}{m}\sum_{i=1}^{m}\left|2M\left\|h_1-h_2\right\|_\infty\right| \\ &= 2M\left\|h_1-h_2\right\|_\infty\end{aligned}$$

综上两项放缩结果可得，

$$|L_S(h_1)-L_S(h_2)|\leq 4M\left\|h_1-h_2\right\|_\infty$$

2.

$$\begin{aligned}\Pr_{S\sim\mathcal{D}^m}\left[\sup_{h\in\mathcal{H}}|L_S(h)|\geqslant\epsilon\right] &= \Pr_{S\sim\mathcal{D}^m}\left[\max_{i\in[k]}\sup_{h\in\mathcal{B}_i}|L_S(h)|\geq\epsilon\right] \\ &\text{上式：对所有的}\mathcal{B}_i\text{，找 h 可让 }|L_S(h)|\text{ 最大，且最大值}\geq\epsilon \\ &\leq \Pr_{S\sim\mathcal{D}^m}\left[\exists i\in[k],\sup_{h\in\mathcal{B}_i}|L_S(h)|\geq\epsilon\right] \\ &\text{上式：只要存在一个 }\mathcal{B}_i\text{，找其中的 h 可让 }|L_S(h)|\text{ 最大，且最大值}\geq\epsilon \\ &= \Pr_{S\sim\mathcal{D}^m}\left[\cup_{i=1}^{k}\sup_{h\in\mathcal{B}_i}|L_S(h)|\geq\epsilon\right] \\ &\leq \sum_{i=1}^{m}\Pr_{S\sim\mathcal{D}^m}\left[\sup_{h\in\mathcal{B}_i}|L_S(h)|\geq\epsilon\right]\end{aligned}$$

3. **3.1** 证明：

由于上一问结论涉及两个 h_i 之间关系，且转化到覆盖涉及 $\|h-h_i\|_\infty$ 之间的关系。所以需要把题目条件进行拆分，通过增添项达到多个 h_i 之间的关系

$$\begin{aligned}\sup_{h\in\mathcal{B}_i}|L_S(h)| &= \sup_{h\in\mathcal{B}_i}|L_S(h)-L_S(h_i)+L_S(h_i)| \\ &\leq \sup_{h\in\mathcal{B}_i}|L_S(h)-L_S(h_i)|+|L_S(h_i)| \\ &\leq 4M\left\|h-h_i\right\|_\infty+|L_S(h_i)|\end{aligned}$$

此时就可以使用覆盖数的定义：由 $\mathcal{N}(\mathcal{H},\frac{\epsilon}{8M})=k,\forall h\in\mathcal{H},\in[k],\|h-h_i\|_\infty\leq\frac{\epsilon}{8M}$ 所以有

$$\begin{aligned}\sup_{h\in\mathcal{B}_i}|L_S(h)| &\leq 4M\frac{\epsilon}{8M}+|L_S(h_i)| \\ &= \frac{\epsilon}{2}+|L_S(h_i)|\end{aligned}$$

综上

$$\begin{aligned}\Pr_{S\sim\mathcal{D}^m}\left[\sup_{h\in\mathcal{B}_i}|L_S(h)|\geqslant\epsilon\right] &\leq \Pr_{S\sim\mathcal{D}^m}\left[|L_S(h_i)|+\frac{\epsilon}{2}\geq\epsilon\right] \\ &= \Pr_{S\sim\mathcal{D}^m}\left[|L_S(h_i)|\geq\frac{\epsilon}{2}\right]\end{aligned}$$

3.2 证明：

$$\begin{aligned}
\Pr_{S\sim\mathcal{D}^m}\left[\sup_{h\in\mathcal{H}}\left|R(h)-\hat{R}_S(h)\right|\geqslant\epsilon\right] &= \Pr_{S\sim\mathcal{D}^m}\left[\sup_{h\in\mathcal{H}}|L_S(h)|\geqslant\epsilon\right] \\
&\leq \sum_{i=1}^{k}\Pr_{S\sim\mathcal{D}^m}\left[\sup_{h\in\mathcal{B}_i}|L_S(h)|\geqslant\epsilon\right] \\
&\text{上式用第二问结论} \\
&\leq \sum_{i=1}^{k}\Pr_{S\sim\mathcal{D}^m}\left[|L_S(h_i)|\geqslant\frac{\epsilon}{2}\right] \\
&\text{上式用 3.1 结论} \\
&= \mathrm{N}(\mathcal{H},\frac{\epsilon}{8M})\cdot\Pr_{S\sim\mathcal{D}^m}\left[|L_S(h_i)|\geqslant\frac{\epsilon}{2}\right] \\
&= \mathrm{N}(\mathcal{H},\frac{\epsilon}{8M})\cdot\Pr_{S\sim\mathcal{D}^m}\left[\left|R(h_i)-\hat{R}_S(h_i)\right|\geqslant\frac{\epsilon}{2}\right]
\end{aligned}$$

其中

$$\begin{aligned}
\Pr_{S\sim\mathcal{D}^m}\left[\left|R(h_i)-\hat{R}_S(h_i)\right|\geqslant\frac{\epsilon}{2}\right] &= \Pr_{S\sim\mathcal{D}^m}\left[\left|\mathbb{E}_{(x,y)\sim\mathcal{D}}[(h_i(x)-y)^2]-\frac{1}{m}\sum_{i=1}^{m}(h(x_i)-y_i)^2\right|\geq\frac{\epsilon}{2}\right] \\
&= \Pr_{S\sim\mathcal{D}^m}\left[\left|\mathbb{E}_{(x,y)\sim\mathcal{D}}[\frac{(h_i(x)-y)^2}{M^2}]-\frac{1}{m}\sum_{i=1}^{m}\frac{(h(x_i)-y_i)^2}{M^2}\right|\geq\frac{\epsilon}{2M^2}\right] \\
&\text{上式为了满足 Hoeffding 前提，对各“变量”进行归一化处理} \\
&\leq 2\cdot exp(-2m\frac{\epsilon^2}{4M^4}) \\
&= 2\cdot exp(-\frac{m\epsilon^2}{2M^4})
\end{aligned}$$

所以综合以上化简式，可得

$$\Pr_{S\sim\mathcal{D}^m}\left[\sup_{h\in\mathcal{H}}\left|R(h)-\hat{R}_S(h)\right|\geqslant\epsilon\right]\leq\mathrm{N}(\mathcal{H},\frac{\epsilon}{8M})\cdot 2e^{-\frac{m\epsilon^2}{2M^4}}$$

三. (40 points) **强化学习 II：基于模型的强化学习**

在强化学习中，“模型”通常指智能体所交互的环境模型。无模型的强化学习根据智能体与环境交互采样到的数据直接进行价值估计或者策略提升，例如 Sarsa 和 Q-learning 算法。而基于模型的强化学习则显式地对“模型”——也就是环境的状态转移函数和奖励函数进行建模，然后智能体就可以额外和学得的环境模型进行交互，对真实环境中样本的需求量往往就会减少，通常会比无模型强化学习算法具有更低的样本复杂度，而且学得的环境模型往往可以进行迁移。

1. (20 points) 考虑一个真实环境转移为 $\mathcal{P}$ 的 MDP：$\mathcal{M}=\langle\mathcal{S},\mathcal{A},\mathcal{R},\mathcal{P},d_0,\gamma\rangle$，若对其学得一个近似的环境转移模型 $\widehat{\mathcal{P}}$ 从而得到 $\widehat{\mathcal{M}}=\langle\mathcal{S},\mathcal{A},\mathcal{R},\widehat{\mathcal{P}},d_0,\gamma\rangle$，且学得的环境模型转移函数 $\widehat{\mathcal{P}}$ 的错误程度满足 $\forall(\boldsymbol{s},\boldsymbol{a})$，$\mathbb{D}_{TV}\left(\widehat{\mathcal{P}}(\cdot\mid\boldsymbol{s},\boldsymbol{a}),\mathcal{P}(\cdot\mid\boldsymbol{s},\boldsymbol{a})\right)\leq\alpha$，并令 $R_{\max}=\max_{(s,a)}|\mathcal{R}(\boldsymbol{s},\boldsymbol{a})|$。MDP 中的 $d_0(\cdot)$ 表示初始状态 s_0 服从的分布。请证明，对任意策略 π，有

$$\left|J\left(\pi,\widehat{\mathcal{M}}\right)-J\left(\pi,\mathcal{M}\right)\right|\leq\frac{2\gamma\alpha R_{\max}}{(1-\gamma)^2} \tag{14}$$

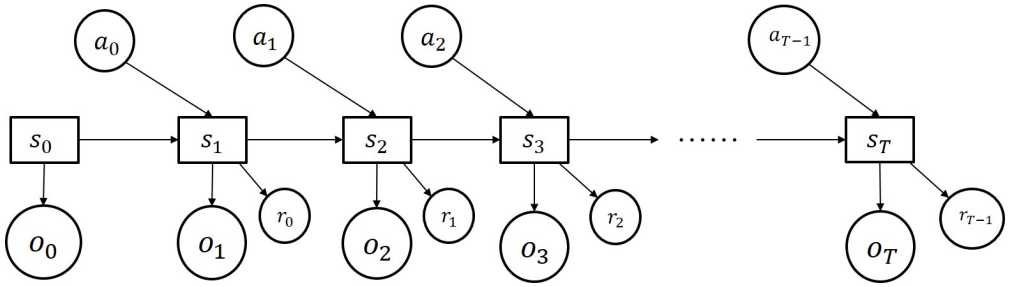
其中 $J(\pi, \mathcal{M})$ 表示某策略 π 在 MDP $\mathcal{M}$ 上交互所得到的强化学习累积折扣回报（目标函数），γ 为折扣系数，$\mathbb{D}_{TV}(\cdot,\cdot)$ 为总变差，皆如上次作业第三题所述。

提示：

(1) 很容易得到强化学习目标函数与值函数满足以下关系：$J(\pi, \mathcal{M}) = \mathbb{E}_{s_0 \sim d_0(s_0)}\left[V^{\pi}_{\mathcal{M}}(s_0)\right]$

(2) 你可以使用上次作业第三题证明的一些性质，或者性质的一般形式。

2. (20 points) 我们一般会用智能体与环境交互从而收集到一些状态动作奖赏数据来学习环境模型。有时，收集到的“状态”数据只是智能体的“观测”，它无法包含全部状态信息，例如交互环境只能提供图像等视觉输入的观测时。若将收集到的观测数据 o 视作证据变量，然后定义隐变量 z 表示智能体的真实工作状态，观测是由隐变量生成。那么我们可以用变分推断方法来学习环境模型，可以画出下面的概率图模型结构：



我们能收集到的数据是轨迹中每一步的观测 $o_{0:T}$ 和动作 $a_{0:T-1}$（这里我们忽略奖赏）。引入如下变分分布：

$$q\left(s_{0:T} \mid o_{0:T}, a_{0:T-1}\right) := \prod_{t=0}^{T} q\left(s_t \mid o_{\leq t}, a_{<t}\right) \tag{15}$$

请你证明，给定动作 $a_{t=0}^{T-1}$ 下观测 $o_{t=0}^{T}$ 的对数似然 $\ln p\left(o_{0:T} \mid a_{0:T-1}\right)$ 的变分下界为

$$\sum_{t=0}^{T}\left[\mathbb{E}_{z_t \sim q\left(\cdot \mid o_{\leq t}, a_{<t}\right)}\left[\ln p\left(o_t \mid z_t\right)\right] - \mathbb{E}_{z_{t-1} \sim q\left(\cdot \mid o_{\leq t-1}, a_{<t-1}\right)}\left[\mathbb{D}_{KL}\left[q\left(\cdot \mid o_{\leq t}, a_{<t}\right) \| p\left(\cdot \mid z_{t-1}, a_{t-1}\right)\right]\right]\right] \tag{16}$$

解：

1. 先对上次作业相关符号定义进行一般化约定，以适用于特定的马尔科夫随机过程 $\mathcal{M}$

   某时刻状态概率 $P^{\pi,\mathcal{M}}(s)$：在特定的马尔科夫随机过程 $\mathcal{M}$ 中，采用策略 π，使 t 时刻 agent 状态为 s 的概率

   占用度量 $\rho^{\pi,\mathcal{M}}(s,a)$：在特定的马尔科夫随机过程 $\mathcal{M}$ 中，状态动作对 (s,a) 被访问到的折扣概率

   $$\rho^{\pi,\mathcal{M}}(s,a) = (1-\gamma)\sum_{t=0}^{\infty}\gamma^t P^{\pi,\mathcal{M}}(s)\cdot\pi(a|s)$$

   使用上次作业最后一题的结论可得

   $$\left|J(\pi,\hat{\mathcal{M}}) - J(\pi,\mathcal{M})\right| \leq \frac{2R_{max}}{1-\gamma}\mathbb{D}_{TV}\left(\rho^{\pi,\hat{\mathcal{M}}}(s,a), \rho^{\pi,\mathcal{M}}(s,a)\right)$$

   对比和本题结论之间的差异，以下需要证明

   $$\mathbb{D}_{TV}\left(\rho^{\pi,\hat{\mathcal{M}}}(s,a), \rho^{\pi,\mathcal{M}}(s,a)\right) \leq \frac{\gamma\alpha}{1-\gamma}$$

$$
\begin{aligned}
2\mathbb{D}_{TV}\left(\rho^{\pi,\hat{\mathcal{M}}}(s,a),\rho^{\pi,\mathcal{M}}(s,a)\right) &= \int_s\int_a\left|\rho^{\pi,\hat{\mathcal{M}}}(s,a)-\rho^{\pi,\mathcal{M}}(s,a)\right| \\
&= \int_s\int_a\left|(1-\gamma)\sum_{t=0}^{\infty}\gamma^t\pi(a|s)\left(P_t^{\pi,\hat{\mathcal{M}}}(s)-P_t^{\pi,\mathcal{M}}(s)\right)\right| \\
&\le \int_a|\pi(a|s)|\int_s(1-\gamma)\sum_{t=0}^{\infty}\gamma^t\left|P_t^{\pi,\hat{\mathcal{M}}}(s)-P_t^{\pi,\mathcal{M}}(s)\right|dsda \\
&= (1-\gamma)\sum_{t=0}^{\infty}\gamma^t\int_s\left|P_t^{\pi,\hat{\mathcal{M}}}(s)-P_t^{\pi,\mathcal{M}}(s)\right|ds
\end{aligned}
$$

注意到 α 和转移函数 $\mathcal{P}(s'|s,a)$ 有关，而转移函数需要确定上一个状态 s' 以及两状态之间的动作 a。所以以下对某时刻状态概率 P(s) 进行转化，使其引入上一时刻状态 s' 和动作 a

$$
P_t^{\pi,\hat{\mathcal{M}}}(s)=\int_a\int_{s'}P_{t-1}^{\pi,\hat{\mathcal{M}}}(s')\cdot\pi(a|s')\cdot\hat{\mathcal{P}}(s|s',a)ds'da
$$

经过对某时刻状态概率 P(s) 的转化，有如下化简

$$
\begin{aligned}
\int_s\left|P_t^{\pi,\hat{\mathcal{M}}}(s)-P_t^{\pi,\mathcal{M}}(s)\right|ds &= \int_s\int_a\int_{s'}\left|P_{t-1}^{\pi,\hat{\mathcal{M}}}(s')\cdot\pi(a|s')\cdot\hat{\mathcal{P}}(s|s',a)-P_{t-1}^{\pi,\mathcal{M}}(s')\cdot\pi(a|s')\cdot\mathcal{P}(s|s',a)\right|ds'dads \\
&= \int_s\int_a\int_{s'}\left|P_{t-1}^{\pi,\hat{\mathcal{M}}}(s')\cdot\pi(a|s')\cdot[\hat{\mathcal{P}}(s|s',a)-\mathcal{P}(s|s',a)]\right| \\
&\qquad+\int_s\int_a\int_{s'}\left|\pi(a|s')\cdot\mathcal{P}(s|s',a)\cdot[P_{t-1}^{\pi,\hat{\mathcal{M}}}(s')-P_{t-1}^{\pi,\mathcal{M}}(s')]\right|ds'dads \\
&\le \int_s\int_a\int_{s'}P_{t-1}^{\pi,\hat{\mathcal{M}}}(s')\cdot\pi(a|s')\cdot\left|\hat{\mathcal{P}}(s|s',a)-\mathcal{P}(s|s',a)\right|ds'dads \\
&\qquad+\int_s\int_a\int_{s'}\pi(a|s')\cdot\mathcal{P}(s|s',a)\cdot\left|P_{t-1}^{\pi,\hat{\mathcal{M}}}(s')-P_{t-1}^{\pi,\mathcal{M}}(s')\right|ds'dads \\
&\le 2\alpha\int_{s'}P_{t-1}^{\pi,\hat{\mathcal{M}}}(s')\int_a\pi(a|s')dads' \\
&\qquad+\int_{s'}\left|P_{t-1}^{\pi,\hat{\mathcal{M}}}(s')-P_{t-1}^{\pi,\mathcal{M}}(s')\right|\cdot\int_a\pi(a|s')\cdot\int_s\mathcal{P}(s|s',a)dsdads' \\
&= 2\alpha+\int_{s'}\left|P_{t-1}^{\pi,\hat{\mathcal{M}}}(s')-P_{t-1}^{\pi,\mathcal{M}}(s')\right|ds'
\end{aligned}
$$

综上，有递推式

$$
\int_s\left|P_t^{\pi,\hat{\mathcal{M}}}(s)-P_t^{\pi,\mathcal{M}}(s)\right|ds\le 2\alpha+\int_{s'}\left|P_{t-1}^{\pi,\hat{\mathcal{M}}}(s')-P_{t-1}^{\pi,\mathcal{M}}(s')\right|ds'
$$

令 $A_t=\int_s\left|P_t^{\pi,\hat{\mathcal{M}}}(s)-P_t^{\pi,\mathcal{M}}(s)\right|ds$ ，所以有递推式 $A_t\le 2\alpha+A_{t-1}$

由于 $A_0=\left|P_0^{\pi,\hat{\mathcal{M}}}(s)-P_0^{\pi,\mathcal{M}}(s)\right|ds=\int_s|0-0|ds=0$

所以有 $A_t \leq 2\alpha t$, 带入原式有如下结论

$$\begin{aligned}
2\mathbb{D}_{TV}(\rho^{\pi,\hat{\mathcal{M}}}(s,a), \rho^{\pi,\mathcal{M}}(s,a)) &= (1-\gamma)\sum_{t=0}^{\infty}\int_s \left|P_t^{\pi,\hat{\mathcal{M}}}(s) - P_t^{\pi,\mathcal{M}}(s)\right| ds \\
&\leq (1-\gamma)\sum_{t=0}^{\infty} 2\alpha t \\
&= (1-\gamma)2\alpha\sum_{t=0}^{\infty} t\gamma^t \\
&\text{等差等比复合数列：使用错位相减即可求出} \\
&= \frac{2\alpha}{1-\gamma}
\end{aligned}$$

综上，带入上次作业的结论中即可得

$$\begin{aligned}
\left|J(\pi,\hat{\mathcal{M}}) - J(\pi,\mathcal{M})\right| &\leq \frac{2R_{max}}{1-\gamma}\mathbb{D}_{TV}\ \left(\rho^{\pi,\hat{\mathcal{M}}}(s,a), \rho^{\pi,\mathcal{M}}(s,a)\right) \\
&\leq \frac{2R_{max}}{1-\gamma}\frac{\gamma\alpha}{1-\gamma} \\
&= \frac{2\gamma\alpha R_{max}}{(1-\gamma)^2}
\end{aligned}$$

2. 由变分推断中真实分布和近似分布之间的关系，

$$lnq(x) = \mathcal{L}(q(z)) + KL(q(z)||p(z|x))$$

带入本题情景中，可以得到

$$lnp(O_{0:T}|a_{0:T-1}) = \mathcal{L}\left(q(Z_{0:T}|O_{0:T}, a_{0:T-1})\right) + \mathbb{D}_{KL}(q(Z_{0:T}|O_{0:T,a_{0:T-1}})||p(Z_{0:T}|O_{0:T,a_{0:T-1}}))$$

其中

$$\begin{aligned}
\mathcal{L}\left(q(Z_{0:T}|O_{0:T}, a_{0:T-1})\right) &= \int_{Z_{0:T}} q(Z_{0:T}|O_{0:T}, a_{0:T-1}) \cdot ln\left(\frac{p(Z_{0:T}, O_{0:T}|a_{0:T-1})}{q(Z_{0:T}|O_{0:T}, a_{0:T-1})}\right) dZ_{0:T} \\
&= \mathbb{E}_{Z_{0:T}\sim q^{T+1}}\left[ln(\frac{p(Z_{0:T}, O_{0:T}|a_{0:T-1})}{q(Z_{0:T}|O_{0:T}, a_{0:T-1})})\right] \\
&= \mathbb{E}_{Z_{0:T}\sim q^{T+1}}\left[ln(\frac{p(O_{0:T}|Z_{0:T}, a_{0:T-1})\cdot p(Z_{0:T|a_{0:T-1}})}{q(Z_{0:T}|O_{0:T}, a_{0:T-1})})\right] \\
&= \mathbb{E}_{Z_{0:T}\sim q^{T+1}}[lnp(O_{0:T}|Z_{0:T}, a_{0:T-1})] \\
&\quad + \mathbb{E}_{Z_{0:T}\sim q^{T+1}}\left[lnp(Z_{0:T|a_{0:T-1}}) - lnq(Z_{0:T}|O_{0:T}, a_{0:T-1})\right]
\end{aligned}$$

以上各个分式可以化简，化简原则如下

$$[1] q(Z_{0:T}|O_{0:T}, a_{0:T-1}) = \prod_{t=0}^{T} q(Z_t|O_{0:t}, a_{0:t-1}) \quad \text{变分分布化简}$$

$$[2] p(O_{0:T|Z_{0:T}, a_{0:T-1}}) = \prod_{t=0}^{T} p(O_t|Z_t, a_{t-1}) \quad \text{由概率图中马尔科夫性可得}$$

使用两个式子进行化简可以得到，

$$\mathcal{L}\left(q(Z_{0:T}|O_{0:T},a_{0:T-1})\right)=\sum_{t=0}^{T}\left[\mathbb{E}_{Z_t\sim q(Z_t|O_{0:t},a_{0:t-1})}[lnp(O_t|Z_t)]\right]$$
$$+\sum_{t=0}^{T}\left[\mathbb{E}_{Z_t\sim q(Z_t|O_{0:t},a_{0:t-1})}[lnp(Z_t|a_{t-1})-lnq(Z_t|O_{0:t},a_{0:t-1}|)]\right]$$

第一个式子: 第一个式子和结论中第一项完全相同。

第二个式子: 为了对 Z_{t-1} 的函数 $\mathbb{D}_{KL}$ 求期望，需要从上式只有 Z_t 中产生 Z_{t-1}，也就需要从概率图模型中发掘 Z_t 和 Z_{t-1} 之间的关系。观察概率图模型可以得到如下等式

$$lnp(Z_t|a_{t-1})=\mathbb{E}_{Z_{t-1}\sim q(Z_{t-1}|O_{0:t-1},a_{0:t-2})}[lnp(Z_t|Z_{t-1},a_{t-1})]$$

所以带入第二个式子中进行化简，构造 Z_{t-1}

$$\begin{aligned}
&\mathbb{E}_{Z_t\sim q(Z_t|O_{0:t},a_{t-1})}[lnp(Z_t|a_{t-1})-lnq(Z_t|O_{0:t},a_{0:t-1})]\\
&\quad=\mathbb{E}_{Z_t\sim q(Z_t|O_{0:t},a_{t-1})}\left[\mathbb{E}_{Z_{t-1}\sim q(Z_{t-1}|O_{0:t-1},a_{t-2})}[lnp(Z_t|Z_{t-1},a_{t-1})]-lnq(Z_t|O_{0:t},a_{0:t-1})\right]\\
&\quad=\mathbb{E}_{Z_{t-1}\sim q(Z_{t-1}|O_{0:t-1},a_{t-2})}\left[\mathbb{E}_{Z_t\sim q(Z_t|O_{0:t},a_{t-1})}\left[ln\left(\frac{p(Z_t|Z_{t-1},a_{t-1})}{q(Z_t|O_{0:t},a_{0:t-1})}\right)\right]\right]\\
&\quad=\mathbb{E}_{Z_{t-1}\sim q(Z_{t-1}|O_{0:t-1},a_{t-2})}\left[-\mathbb{D}_{KL}\left(q(Z_t|O_{0:t},a_{0:t-1})||p(Z_t|Z_{t-1},a_{t-1})\right)\right]\\
&\quad=-\mathbb{E}_{Z_{t-1}\sim q(Z_{t-1}|O_{0:t-1},a_{t-2})}\left[\mathbb{D}_{KL}\left(q(Z_t|O_{0:t},a_{0:t-1})||p(Z_t|Z_{t-1},a_{t-1})\right)\right]
\end{aligned}$$

综合第一式和对第二式的化简结果可得, 给定动作 $a_{t=0}^{T-1}$ 下观测 $o_{t=0}^{T}$ 的对数似然 $\ln p\left(o_{0:T}\mid a_{0:T-1}\right)$ 的变分下界 $\mathcal{L}(q)$ 为

$$\begin{aligned}
&\mathcal{L}\left(q(Z_{0:T}|O_{0:T},a_{0:T-1})\right)\\
&\quad=\sum_{t=0}^{T}\left[\mathbb{E}_{z_t\sim q\left(\cdot|o_{\leq t},a_{<t}\right)}\left[\ln p\left(o_t\mid z_t\right)\right]-\mathbb{E}_{z_{t-1}\sim q\left(\cdot|o_{\leq t-1},a_{<t-1}\right)}\left[\mathbb{D}_{KL}\left[q\left(\cdot\mid o_{\leq t},a_{<t}\right)\|p\left(\cdot\mid z_{t-1},a_{t-1}\right)\right]\right]\right]
\end{aligned}$$