基于神经网络的代理模型

神经算子：

$$a(x) \to \int k(x,y)\, a(y)\, dy + b(x)$$

$k(x,y)$ 核函数，可再生希尔伯特空间，Koopman 算子

① 投影到一组基底上

$$a(x) = \sum a_i \psi_i(x) \qquad \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_l \end{pmatrix} = A \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_l \end{pmatrix}$$

$$u(x) = \sum u_i \psi_i(x)$$

PCA-net

给定 Hilbert 空间 $\mathcal{H}$，给定概率测度 ν，有有限 4 阶矩 $E_{u\sim\nu}\|u\|^4 < \infty$，我们有数据 $\{u_i\}_{i=1}^n \sim \nu$

$C_n := \frac{1}{n}\sum_{i=1}^n u_i \otimes u_i$，它的前 k 个特征值，特征向量

满足

$$\lambda_{1,n} \geq \lambda_{2,n} \cdots \geq \lambda_{k,n}, \quad \phi_{1,n}, \phi_{2,n} \cdots \phi_{k,n}$$

我们定义投影算子 $F_{\mathcal{H}} : \mathcal{H} \to R^k$

$$F_{\mathcal{H}}(u) = (\langle u, \phi_{1,n}\rangle, \langle u, \phi_{2,n}\rangle \cdots \langle u, \phi_{k,n}\rangle)^T \in R^k$$

和 PCA decoder $G_{\mathcal{H}} : R^k \to \mathcal{H}$

$$G_{\mathcal{H}}(s) = \sum_{i=1}^k s_i \phi_{i,n}$$

投影算子：$\Pi_{V_{k,n}} = \sum_{i=1}^k \phi_{i,n} \otimes \phi_{i,n} : \mathcal{H} \to V_{k,n}$

$$V_{k,n} = \mathrm{span}\{\phi_{1n}, \phi_{2n} \cdots \phi_{kn}\} \subset \mathcal{H}$$

PCA-net 近似误差

$$\mathbb{E}_{\{a_j\}\sim\nu}\, \mathbb{E}_{a\sim\nu} \| G^+(a) - G_U \circ f \circ F_A(a) \|_U^2$$

$$= 2\mathbb{E}_{\{a_j\}\sim\nu}\, \mathbb{E}_{a\sim\nu} \Big(\underbrace{\| G^+(a) - G_U \circ f^+ \circ F_A(a) \|_U^2}_{\text{投影误差}} + \underbrace{\| G_U \circ (f^+ - f) F_A(a) \|_U^2}_{\text{神经网络近似误差}} \Big)$$

$$f^+ = F_U \circ G^+ \circ G_A : R^k \to R^k$$

对于投影误差：我们定义 PCA 投影

$$C = \mathbb{E}_{u\sim\nu} [u \otimes u] \qquad \operatorname{tr} C = \mathbb{E}_{u\sim\nu} \|u\|^2 < \infty$$

$$C = \sum_{i=1}^{\infty} \lambda_i \phi_i \otimes \phi_i$$

$$\min_{V\in\mathcal{V}} R(V) = \mathbb{E}_{u\sim\nu} \| u - \Pi_V u \|^2$$，其中 Π_V 是投影

到 k 维子空间 V 中，$V = \operatorname{span}\{\phi_1, \phi_2 \cdots \phi_k\}$

我们希望用 $\lambda_{i,n} \approx \lambda_i \qquad \phi_{i,n} \approx \phi_i \qquad F_A^+ \quad G_U^+$

$$\mathbb{E}\,\mathbb{E}\,\|G^{+}(a)-G_{\mathcal{U}}\circ f^{+}\circ F_A(a)\|_{\mathcal{U}}^2$$

$$= \mathbb{E}\,\mathbb{E}\left(\|G^{+}(a)-G_{\mathcal{U}}\circ F_{\mathcal{U}}\circ G^{+}\circ G_A\circ F_A(a)\|_{\mathcal{U}}^2\right)$$

$$= \mathbb{E}\,\mathbb{E}\,\|G^{+}(a)-\Pi_{\mathcal{U}_{k,n}}\circ G^{+}\circ \Pi_{A_{k,n}}(a)\|_{\mathcal{U}}^2$$

$$\leq 2\,\mathbb{E}\,\mathbb{E}\,\|G^{+}(a)-\Pi_{\mathcal{U}_{k,n}}\circ G^{+}(a)\|_{\mathcal{U}}^2+\|\Pi_{\mathcal{U}_{k,n}}\circ G^{+}(a)-\Pi_{\mathcal{U}_{k,n}}\circ G^{+}\circ \Pi_{A_{k,n}}(a)\|_{\mathcal{U}}^2$$

第一项满足

$$\mathbb{E}_{a\sim\nu}\,\|G^{+}(a)-\Pi_{\mathcal{U}_{k,n}}\circ G^{+}(a)\|_{\mathcal{U}}^2=\mathbb{E}_{u\sim G^{+}_{\#}\nu}\,\|u-\Pi_{\mathcal{U}_{k,n}}(u)\|_{\mathcal{U}}^2$$

假设 G^{+} 满足全局 L-Lipschitz

$$\|G^{+}(a)-G^{+}(a')\|_{\mathcal{U}}\leq L\,\|a-a'\|_A$$

第二项满足

$$\|\Pi_{\mathcal{U}_{k,n}}\circ G^{+}(a)-\Pi_{\mathcal{U}_{k,n}}\circ G^{+}\circ \Pi_{A_{k,n}}(a)\|_{\mathcal{U}}^2\leq \|G^{+}\circ(a)-G^{+}\circ \Pi_{A_{k,n}}(a)\|_{\mathcal{U}}^2$$

$$\leq L^2\,\|a-\Pi_{A_{k,n}}(a)\|_A^2$$

引理：$$\mathbb{E}_{\{a_j\}\sim\nu}\,\mathbb{E}_{a\sim\nu}\,\|a-\Pi_{A_{k,n}}(a)\|_A^2\leq \mathbb{E}_{\{a_j\}\sim\nu}\,\mathbb{E}_{a\sim\nu}\,\|a-\Pi_{A_k}(a)\|_A^2+C\sqrt{\frac{k}{n}}$$

其中 n 是数据量，k是低维空间维度

记 $\mathbb{E}_{a\sim\nu}\|a-\Pi_A a\|_A^2=R(A)$，$R_n(A)=\frac{1}{n}\sum_{j=1}^{n}\|a_j-\Pi_A a_j\|^2$

我们有 $$R_n(A_{k,n})=\sum_{j=k+1}^{n}\lambda_{j,n}$$

$$R(A_{k,n}) = tr(C) - \mathbb{E}_{a\sim\nu}\langle \Pi_{A_{k,n}} a, \Pi_{A_{k,n}} a\rangle_A$$

$$R_n(A_{k,n}) = tr(C_n) - \frac{1}{n}\sum_j \langle \Pi_{A_{k,n}} a_j - \Pi_{A_{k,n}} a_j \rangle$$

$$\mathbb{E}_{\{a_j\}\sim\nu}[R(A_{k,n}) - R_n(A_{k,n})]$$

$$\langle \Pi, u\otimes v\rangle_{HS} = \langle \Pi u, \Pi v\rangle_A$$

$$= \mathbb{E}_{\{a_j\}\sim\nu} \langle \Pi_{A_{k,n}}, C - C_n\rangle_{HS} \leq \mathbb{E}_{\{a_j\}\sim\nu} \underbrace{\|\Pi_{A_{k,n}}\|_{HS}}_{\sqrt{k}} \underbrace{\|C-C_n\|_{HS}}_{\sqrt{\frac{1}{n}}}$$

$$\leq C\sqrt{\frac{k}{n}}$$

$$\mathbb{E}_{\{a_j\}\sim\nu}\|C-C_n\|_{HS} \approx \sqrt{\frac{C}{n}}$$

$$\mathbb{E}_{\{a_j\}\sim\nu} R(A_{k,n}) \leq C\sqrt{\frac{k}{n}} + \sum_{j=k+1}^{n} \lambda_{j,n}$$

引理 $\mathbb{E}_{\{a_j\}\sim\nu} \sum_{j=k+1}^{n} \lambda_{j,n} \leq \sum_{j=k+1}^{\infty} \lambda_j = R(A_k)$

对待神经网络近似误差

$$\mathbb{E}_{\{a_j\}\sim\nu}\mathbb{E}_{a\sim\nu} \|G_u\circ(F_u G^+ G_A - f) F_A\|_u^2$$

我们极小化 $\|F_u G^+ G_A(a) - f(a)\|_2^2$

为了使用万有逼近定理：我们要验证

i) 连续性 (Lipschitz)

ii) 在紧集上（只考虑 $a\in[-M,M]^d$）

$$\|F_u G^+ G_A(a) - F_u G^+ G_A(b)\|_2^2$$

我们只需要 F_u, G_A 是 Lipschitz

$$\|F_u(u) - F_u(v)\|_2^2 = \sum \langle u-v, \phi_j \rangle_u^2$$

$$\le \|u-v\|_u^2 \quad \text{(Parseval)}$$

$$\|G_A(a) - G_A(b)\|_A^2 = \|\sum (a_i - b_i)\psi_j\|_A^2$$

$$= \|a-b\|_2^2$$

那么我们有 $\|F_u G^+ G_A(a) - f(a)\|_2^2 < \varepsilon$

$$\forall \ a \in [-M, M]^d$$

最后：

$$E_{a\sim\nu} \|G_u \circ (F_u G^+ G_A - f) F_A\|_u^2$$

$$= \int_{F_A(a)\in[-M,M]^d} \|G_u \circ (\underbrace{F_u G^+ G_A - f}_{\le \varepsilon^2}) \circ F_A(a)\|_u^2 d\mu(a) \quad G_u \text{ 是 Lipschitz}$$

$$+ \int_{F_A(a)\notin[-M,M]^d} \|G_u \circ (F_u G^+ G_A - f) F_A(a)\|_u^2 d\mu(a)$$

$$\le \int_{F_A(a)\notin[-M,M]^d} \|(F_u G^+ G_A - f) F_A(a)\|^2 d\mu(a)$$

f 在 $[-M, M]^d$ 以外为 0

$$= \int_{F_A(a)\notin[-M,M]^d} \|F_u G^+ G_A F_A(a)\|^2 d\mu(a)$$

由于 $F_u G^+ G_A$ 是 L-Lipschitz $\quad |F_A(a)|_2 \le \|a\|_A$

$$\le \int_{F_A(a)\notin[-M,M]^d} \left(|F_u G^+ G_A(0)|_2 + L\|a\|_A\right)^2 d\mu(a)$$

$$\le 2\mu\left(F_A(a)\notin[-M,M]^d\right) |F_u G^+ G_A(0)|_2^2$$

$$+ L^2 \mu^2 \mathop{E}_{a\sim\nu} \|a\|^4$$

当 M 足够大 μ 足够小