Online SPGP-SARSA

Temporal Difference Learning via Recursive Sparse GP Regression

John Martin Jr.

November 17, 2017

Sparse GP Temporal Differences

Rewards are noisy outputs of a residual value process

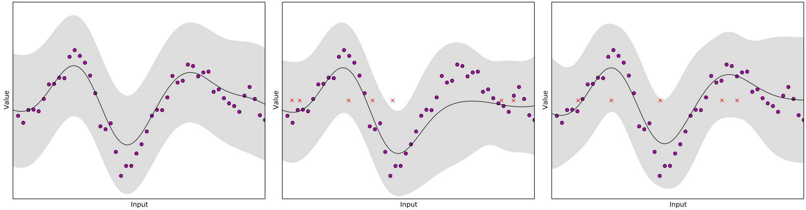
$$\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & -\gamma & \cdots & 0 \\ 0 & 1 & -\gamma & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & 1 & -\gamma \end{pmatrix} \begin{pmatrix} V(\mathbf{x}_1) \\ V(\mathbf{x}_2) \\ \vdots \\ V(\mathbf{x}_t) \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \end{pmatrix}.$$

- Assume $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_{vv}(\mathbf{x}))$, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$
- Expand probability space with M pseudo inputs $\mathbf{z}$
- Collapse by conditioning on $\mathbf{z}$

$$\tilde{v}(\mathbf{x}_t) = \boldsymbol{\alpha}_t^\top \mathbf{r}_t, \qquad \tilde{p}(\mathbf{x}_t) = k(\mathbf{x}_t, \mathbf{x}_t) - \mathbf{k}_k^\top(\mathbf{x}_t)\mathbf{P}_{kk}\mathbf{k}_k(\mathbf{x}_t). \tag{1}$$

- Parameters: $\boldsymbol{\alpha}_t$, $\mathbf{P}_{kk}$ depend only on rank M inverses

Choosing the Pseudo Inputs



Optimization distributes probability mass to explain the data

- Standard GP Regression scales with N^3
- Sparse method scales with NM^2, where $M \ll N$

Deriving and Online Algorithm

- Standard SPGP-SARSA is inherently offline

Theorem

Let $\mathbf{K}_t$ be $t \times t$ symmetric positive definite with partition

$$\mathbf{K}_t = \begin{pmatrix} \mathbf{K}_{t-1} & \mathbf{k}_{t-1}(\mathbf{x}_t) \\ \mathbf{k}_{t-1}(\mathbf{x}_t)^\top & k_{tt} \end{pmatrix}. \tag{2}$$

Define $s_t = k_{tt} - \mathbf{k}_{t-1}^\top(\mathbf{x}_t)\mathbf{K}_{t-1}^{-1}\mathbf{k}_{t-1}(\mathbf{x}_t)$. The inverse is

$$\mathbf{K}_t^{-1} = \begin{pmatrix} \mathbf{K}_{t-1}^{-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} + \frac{1}{s_t} \begin{pmatrix} \mathbf{K}_{t-1}^{-1}\mathbf{k}_{t-1}(\mathbf{x}_t) \\ -1 \end{pmatrix} \begin{pmatrix} \mathbf{k}_{t-1}^\top(\mathbf{x}_t)\mathbf{K}_{t-1}^{-1} & -1 \end{pmatrix}.$$