随机过程

报告副标题

作者 1 [1]、**作者 2** [2]

December 21st, 2024

[1]Tongji University
[2]Department of Computer Science and Technology

Table of Contents

Snelson and Ghahramani proposes the idea of having pseudo data[1] , which is later referred to as Fully independent training condition(FITC).

Key Idea

Augment the training data $(\mathbf{X}, \mathbf{y})$ with pseudo data $\mathbf{u}$ at location $\mathbf{Z}$.

$$p\left(\begin{bmatrix}\mathbf{y}\\ \mathbf{u}\end{bmatrix} \middle| \begin{bmatrix}\mathbf{X}\\ \mathbf{Z}\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix}\mathbf{y}\\ \mathbf{u}\end{bmatrix} \middle| \mathbf{0}, \begin{bmatrix}\mathbf{K_{NN}} + \sigma^2\mathbf{I_N} & \mathbf{K_{NM}}\\ \mathbf{K_{NM}^\top} & \mathbf{K_{MM}}\end{bmatrix}\right)$$

where $\mathbf{K_{NN}} = \mathbf{K}(\mathbf{X}, \mathbf{X}), \mathbf{K_{NM}} = \mathbf{K}(\mathbf{X}, \mathbf{Z}), \mathbf{K_{MM}} = \mathbf{K}(\mathbf{Z}, \mathbf{Z})$ and $\mathbf{Z}$ is the cluster centroids returned by the K-means algorithm. Assuming that $\mathbf{y}$ and $f(\mathbf{X}_*)$, where $\mathbf{X}_*$ are the test inputs, are conditionally independent given $\mathbf{u}$, we have $Cov(f(\mathbf{X}_*), \mathbf{y}) = \mathbf{K}_{*M}\mathbf{K}_{MM}^{-1}\mathbf{K}_{MN}$

- Thanks to the marginalization property of Gaussian distribution,

$$p(\mathbf{y}|\mathbf{X}) = \int_{\mathbf{u}} p(\mathbf{y}, \mathbf{u}|\mathbf{X}, \mathbf{Z})$$

- Further re-arrange the notation:

$$p(\mathbf{y}, \mathbf{u}|\mathbf{X}, \mathbf{Z}) = p(\mathbf{y}|\mathbf{u}, \mathbf{X}, \mathbf{Z})p(\mathbf{u}|\mathbf{X}, \mathbf{Z}) = p(\mathbf{y}|\mathbf{u}, \mathbf{X}, \mathbf{Z})p(\mathbf{u}|\mathbf{Z})$$

where $p(\mathbf{y}|\mathbf{u}, \mathbf{X}, \mathbf{Z}) = \mathcal{N}(\mathbf{y}|\mathbf{K}_{\mathbf{NM}}\mathbf{K}_{\mathbf{MM}}^{-1}\mathbf{u}, \mathbf{K}_{\mathbf{NN}} - \mathbf{K}_{\mathbf{NM}}\mathbf{K}_{\mathbf{MM}}^{-1}\mathbf{K}_{\mathbf{NM}}^{\top}) + \sigma^2\mathbf{I}$,
$p(\mathbf{u}|\mathbf{Z}) = \mathcal{N}(\mathbf{u}|0, \mathbf{K}_{\mathbf{MM}})$

Key Idea

The FITC (Fully Independent Training Conditional) approxiamtion assuames

$$\tilde{p}(\mathbf{y}|\mathbf{u},\mathbf{X},\mathbf{Z}) = \mathcal{N}(\mathbf{y}|\mathbf{K}_{\mathbf{NM}}\mathbf{K}_{\mathbf{MM}}^{-1}\mathbf{u}, \mathit{diag}[\mathbf{K}_{\mathbf{NN}} - \mathbf{Q}_{\mathbf{NN}}] + \sigma^2\mathbf{I}_{\mathbf{N}}),$$

where $\mathbf{Q}_{\mathbf{NN}} = \mathbf{K}_{\mathbf{NM}}\mathbf{K}_{\mathbf{MM}}^{-1}\mathbf{K}_{\mathbf{NM}}^{\top}$

- Marginalize $\mathbf{u}$ from the model definition:

$$\tilde{p}(\mathbf{y}|\mathbf{X},\mathbf{Z}) = \mathcal{N}(\mathbf{y}|0, \mathbf{Q}_{\mathbf{NN}} + \mathit{diag}[\mathbf{K}_{\mathbf{NN}} - \mathbf{Q}_{\mathbf{NN}}] + \sigma^2\mathbf{I}_{\mathbf{N}}) = \mathcal{N}(\mathbf{y}|\mathbf{Q}_{\mathbf{NN}} + \mathbf{D}_{\mathbf{N}}) := \mathcal{N}(\mathbf{y}|\mathbf{K}_*).$$

where $\mathbf{D}_{\mathbf{N}} = \mathit{diag}[\mathbf{K}_{\mathbf{NN}} - \mathbf{Q}_{\mathbf{NN}}] + \sigma^2\mathbf{I}_{\mathbf{N}}$

FITC Approximation

1 Sparse Gaussian Process

- Start by considering the Cholesky decomposition $\mathbf{K_{MM}} = \mathbf{U_M}\mathbf{U_M}^\top$.$({\color{red}\mathcal{O}(M^3)})$
- We have $\mathbf{Q_{NN}} = \mathbf{V_{MN}}^\top\mathbf{V_{MN}}$,where $\mathbf{V_{MN}} = \mathbf{U_M}^{-1}\mathbf{K_{MN}}$.$({\color{red}\mathcal{O}(M^2N)})$
- The log-determinant term $\log(|\mathbf{K}_*|)$ can be rewritten as

$$\log(|\mathbf{V}^\top\mathbf{V} + \mathbf{D_N}|) = \log(|\mathbf{D_N}| \cdot |\mathbf{I_M} + \mathbf{V}\mathbf{D_N}^{-1}\mathbf{V}^\top|) = \sum_i \log(\mathbf{d}_i) + 2\sum_i \log(\mathbf{L}_{ii}),$$

  where $\mathbf{d}_i$ is the ith diagonal terms of $\mathbf{D_N}$, $\mathbf{L_M}\mathbf{L_M}^\top = \mathbf{I_M} + \mathbf{V_{MN}}\mathbf{D_N}^{-1}\mathbf{V_{MN}}^\top$.$({\color{red}\mathcal{O}(M^2N)})$
- Using Woodbury formula, we have:

$$\mathbf{K}_*^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{V}^\top(\mathbf{I_M} + \mathbf{V}\mathbf{D_N}^{-1}\mathbf{V}^\top)^{-1}\mathbf{V}\mathbf{D}^{-1} = \mathbf{D}^{-1} - (\underbrace{\mathbf{L}^{-1}\mathbf{V}\mathbf{D}^{-1}}_{\color{red}\mathcal{O}(M^2N)})^\top(\mathbf{L}^{-1}\mathbf{V}\mathbf{D}^{-1}).$$

Table of Contents

[1] Snelson, Edward and Zoubin Ghahramani. "Sparse Gaussian Processes using Pseudo-inputs." Neural Information Processing Systems (2005).

[2] Hildo Bijl, Jan-Willem van Wingerden, Thomas B. Schön, Michel Verhaegen, Online sparse Gaussian process regression using FITC and PITC approximations

随机过程

Thank you for listening!
Any questions?