Gaussian Process with Pseudo Data

Snelson and Ghahramani [2006] proposes the idea of having pseudo data, which is later referred to as NMlly independent training condition(FITC).

Key Idea

Augment the training data $(\mathbf{X}, \mathbf{y})$ with pseudo data $\mathbf{u}$ at location $\mathbf{Z}$. Then try to use $\mathbf{K_{MM}}$ to approximate $\mathbf{K_{NN}}$

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

where $\mathbf{K_{NN}} = \mathbf{K}(\mathbf{X}, \mathbf{X}), \mathbf{K_{NM}} = \mathbf{K}(\mathbf{X}, \mathbf{Z}), \mathbf{K_{MM}} = \mathbf{K}(\mathbf{Z}, \mathbf{Z})$.

Gaussian Process with Pseudo Data

- 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:

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

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

Key Idea

The FITC approxiamtion assuames

$$\tilde{p}(\mathbf{y}|\mathbf{u}, \mathbf{X}, \mathbf{Z}) = \mathcal{N}(\mathbf{y}|\mathbf{K_{NM}}\mathbf{K_{MM}^{-1}}\mathbf{u}, diag[\mathbf{K_{NN}} - \mathbf{Q_{NN}}] + \sigma^2\mathbf{I_N}),$$

where $\mathbf{Q_{NN}} = \mathbf{K_{NM}}\mathbf{K_{MM}^{-1}}\mathbf{K_{fu}^{\top}}$

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

$$\tilde{p}(\mathbf{y}|\mathbf{X}, \mathbf{Z}) = \mathcal{N}(\mathbf{y}|0, \mathbf{Q_{NN}} + diag[\mathbf{K_{NN}} - \mathbf{Q_{NN}}] + \sigma^2\mathbf{I_N}) = \mathcal{N}(\mathbf{y}|\mathbf{Q_{NN}} + \mathbf{D_N}).$$

where $\mathbf{D_N} = diag[\mathbf{K_{NN}} - \mathbf{Q_{NN}}] + \sigma^2\mathbf{I_N}$

FITC Approximation

- Start by considering the Cholesky decomposition $\mathbf{K_{MM}} = \mathbf{U_M}\mathbf{U_M^\top}$.$(\mathcal{O}(M^3))$
- We have $\mathbf{Q_{NN}} = \mathbf{V_{MN}^\top}\mathbf{V_{MN}}$,where $\mathbf{V_{MN}} = \mathbf{U_M^{-1}}\mathbf{K_{MN}}$.$(\mathcal{O}(M^2N))$
- 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 $\mathbf{D_N}$, $\mathbf{L_M}\mathbf{L_M^\top} = \mathbf{I_M} + \mathbf{V_{MN}}\mathbf{D_N^{-1}}\mathbf{V_{MN}^\top}$.$(\mathcal{O}(M^2N))$

- Using Woodbury formula, we have:

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