Snelson and Ghahramani [2006] proposes the idea of having pseudo data, 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 $\mathbf{Z}$. Then try to use $\mathbf{K}_{\mathbf{uu}}$ to approximate $\mathbf{K}_{\mathbf{ff}}$

$$p\left(\begin{bmatrix}\mathbf{y}\\ \mathbf{u}\end{bmatrix} \middle| \begin{bmatrix}\mathbf{X}\\ \mathbf{Z}\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix}\mathbf{y}\\ \mathbf{u}\end{bmatrix}; \mathbf{0}, \begin{bmatrix}\mathbf{K}_{\mathbf{yy}} & \mathbf{K}_{\mathbf{yu}}\\ \mathbf{K}_{\mathbf{yu}}^\top & \mathbf{K}_{\mathbf{uu}}\end{bmatrix}\right)$$

where $\mathbf{K}_{\mathbf{yy}} = \mathbf{K}(\mathbf{X}, \mathbf{X}), \mathbf{K}_{\mathbf{yu}} = \mathbf{K}(\mathbf{X}, \mathbf{Z}), \mathbf{K}_{\mathbf{uu}} = \mathbf{K}(\mathbf{Z}, \mathbf{Z})$.

- 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 $p(\mathbf{y}|\mathbf{u}, \mathbf{X}, \mathbf{Z}) = \mathcal{N}(\mathbf{y}|\mathbf{K}_{\mathbf{yu}}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{u}, \mathbf{K}_{\mathbf{yy}} - \mathbf{K}_{\mathbf{yu}}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{K}_{\mathbf{yu}}^{\top}), p(\mathbf{u}|\mathbf{Z}) = \mathcal{N}(\mathbf{u}|0, \mathbf{K}_{\mathbf{uu}})$

Key Idea

The FITC approxiamtion assuames

$$\tilde{p}(\mathbf{y}|\mathbf{u}, \mathbf{X}, \mathbf{Z}) = \mathcal{N}(\mathbf{y}|\mathbf{K}_{\mathbf{yu}}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{u}, \Lambda),$$

where $\Lambda = (\mathbf{K}_{\mathbf{yy}} - \mathbf{K}_{\mathbf{yu}}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{K}_{\mathbf{yu}}^{\top}) \circ \mathbf{I}$

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

$$\tilde{p}(\mathbf{y}|\mathbf{X}, \mathbf{Z}) = \mathcal{N}(\mathbf{y}|\mathbf{K}_{\mathbf{yu}}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{K}_{\mathbf{yu}}^{\top} + \Lambda).$$

- Then apply Woodbury formula:

$$(\mathbf{K}_z\mathbf{K}_{zz}^{-1}\mathbf{K}_z^{\top} + \Lambda)^{-1} = \Lambda^{-1} - \Lambda^{-1}\mathbf{K}_z(\mathbf{K}_{zz} + \mathbf{K}_z^{\top}\Lambda^{-1}\mathbf{K}_z)^{-1}\mathbf{K}_z^{\top}\Lambda^{-1}$$