Predictive distributions of wide ensembles for various training methods.

Method	Layers trained	$\mu(\boldsymbol{x})$	$\Sigma(\boldsymbol{x}, \boldsymbol{x}')$
NNGP	Final	$\mathcal{K}_{\boldsymbol{x}\mathcal{X}}(\mathcal{K}_{\mathcal{X}\mathcal{X}}+\sigma^2 I)^{-1}\mathcal{Y}$	$\mathcal{K}_{\boldsymbol{x}\boldsymbol{x}'} - \mathcal{K}_{\boldsymbol{x}\mathcal{X}}(\mathcal{K}_{\mathcal{X}\mathcal{X}}+\sigma^2 I)^{-1}\mathcal{K}_{\mathcal{X}\boldsymbol{x}'}$
Deep Ensembles	All	$\Theta_{\boldsymbol{x}\mathcal{X}}\Theta_{\mathcal{X}\mathcal{X}}^{-1}\mathcal{Y}$	$\mathcal{K}_{\boldsymbol{x}\boldsymbol{x}'} - \left(\Theta_{\boldsymbol{x}\mathcal{X}}\Theta_{\mathcal{X}\mathcal{X}}^{-1}\mathcal{K}_{\mathcal{X}\boldsymbol{x}'}+h.c.\right)$ $\Theta_{\boldsymbol{x}\mathcal{X}}\Theta_{\mathcal{X}\mathcal{X}}^{-1}\mathcal{K}_{\mathcal{X}\mathcal{X}}\Theta_{\mathcal{X}\mathcal{X}}^{-1}\Theta_{\mathcal{X}\boldsymbol{x}'}$
Randomised Prior	All	$\Theta_{\boldsymbol{x}\mathcal{X}}(\Theta_{\mathcal{X}\mathcal{X}}+\sigma^2 I)^{-1}\mathcal{Y}$	$\mathcal{K}_{\boldsymbol{x}\boldsymbol{x}'} - \left(\Theta_{\boldsymbol{x}\mathcal{X}}(\Theta_{\mathcal{X}\mathcal{X}}+\sigma^2 I)^{-1}\mathcal{K}_{\mathcal{X}\boldsymbol{x}'}+h.c.\right)$ $+\Theta_{\boldsymbol{x}\mathcal{X}}(\Theta_{\mathcal{X}\mathcal{X}}+\sigma^2 I)^{-1}(\mathcal{K}_{\mathcal{X}\mathcal{X}}+\sigma^2 I)(\Theta_{\mathcal{X}\mathcal{X}}+\sigma^2 I)^{-1}\Theta_{\mathcal{X}\boldsymbol{x}'}$
NTKGP	All (ours)	$\Theta_{\boldsymbol{x}\mathcal{X}}(\Theta_{\mathcal{X}\mathcal{X}}+\sigma^2 I)^{-1}\mathcal{Y}$	$\Theta_{\boldsymbol{x}\boldsymbol{x}'} - \Theta_{\boldsymbol{x}\mathcal{X}}(\Theta_{\mathcal{X}\mathcal{X}}+\sigma^2 I)^{-1}\Theta_{\mathcal{X}\boldsymbol{x}'}$