Avoiding Pathologies in Very Deep Networks

David Duvenaud, Oren Rippel, Ryan Adams, Zoubin Ghahramani

Nonparametric Priors on Deep Nets

Deep GPs are compositions of functions, each $f^{(\ell)} \stackrel{\text{\tiny ind}}{\sim} \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$.

$$m{f}^{(1:L)}(\mathbf{x}) = m{f}^{(L)}(m{f}^{(L-1)}(\dots m{f}^{(2)}(m{f}^{(1)}(\mathbf{x}))\dots))$$

Can be derived as either

1. neural nets with nonparametric activation functions

2. neural nets with infintely-many parametric hidden nodes

