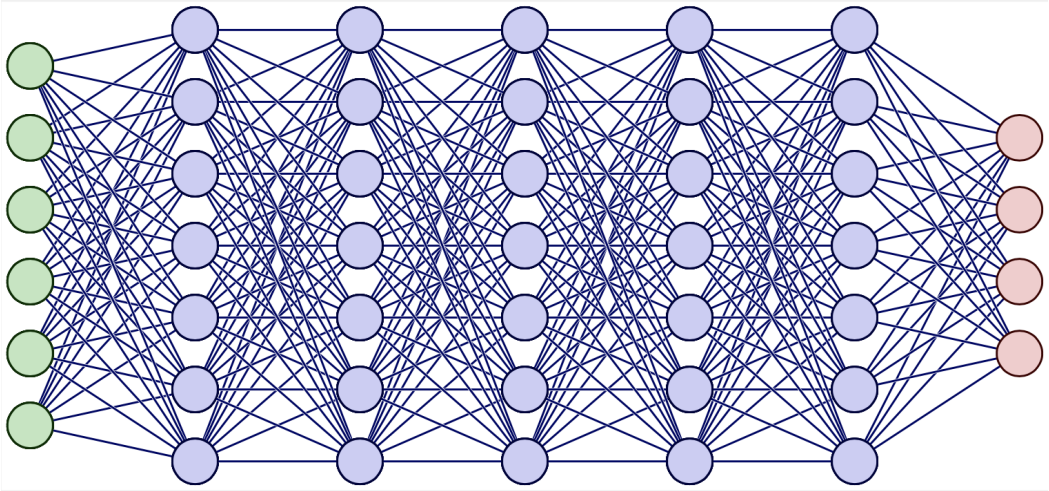
EFT for ML (and flows)



L Del Debbio

Higgs Centre for Theoretical Physics
University of Edinburgh

based on paper [Roberts et al 21]

Bayesian inference

$$y_I = \int dx\, C_I(x) f(x)$$

Bayesian solution:

$$f_i = f(x_i) \quad \longrightarrow \quad p(f_1 \dots f_N | y, \mathcal{H}) = \frac{p(y|f,\mathcal{H})p(f|\mathcal{H})}{p(y|\mathcal{H})}$$

$\hookrightarrow$ dependence on the choice of the prior

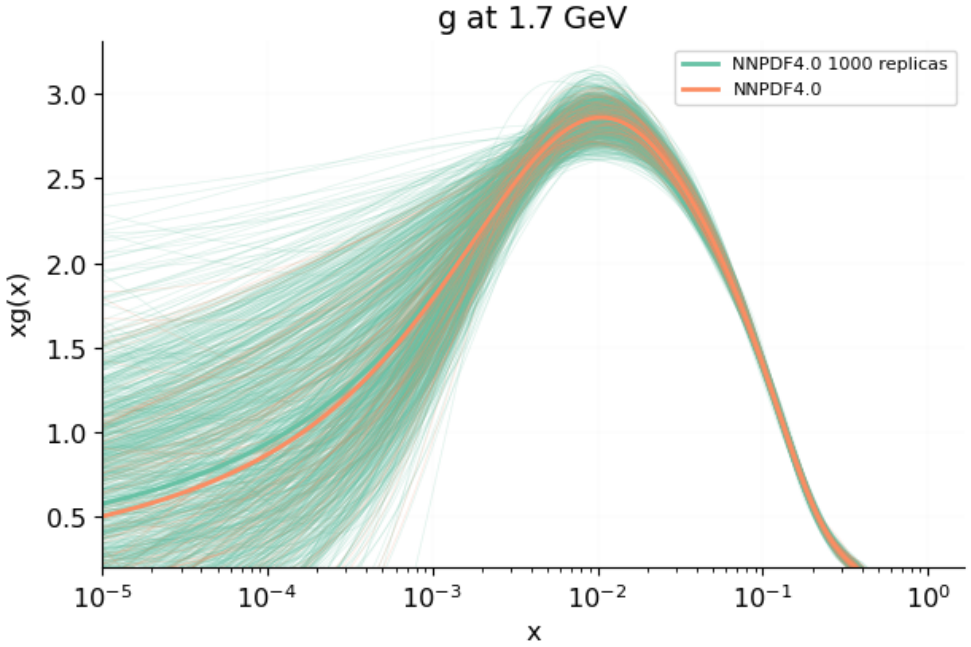
Gaussian processes: $f \sim \mathcal{GP}(m, k)$

$$\mathbb{E}[f_i] = m(x_i)\,, \quad \mathrm{Cov}[f_i, f_j] = k(x_i, x_j)$$

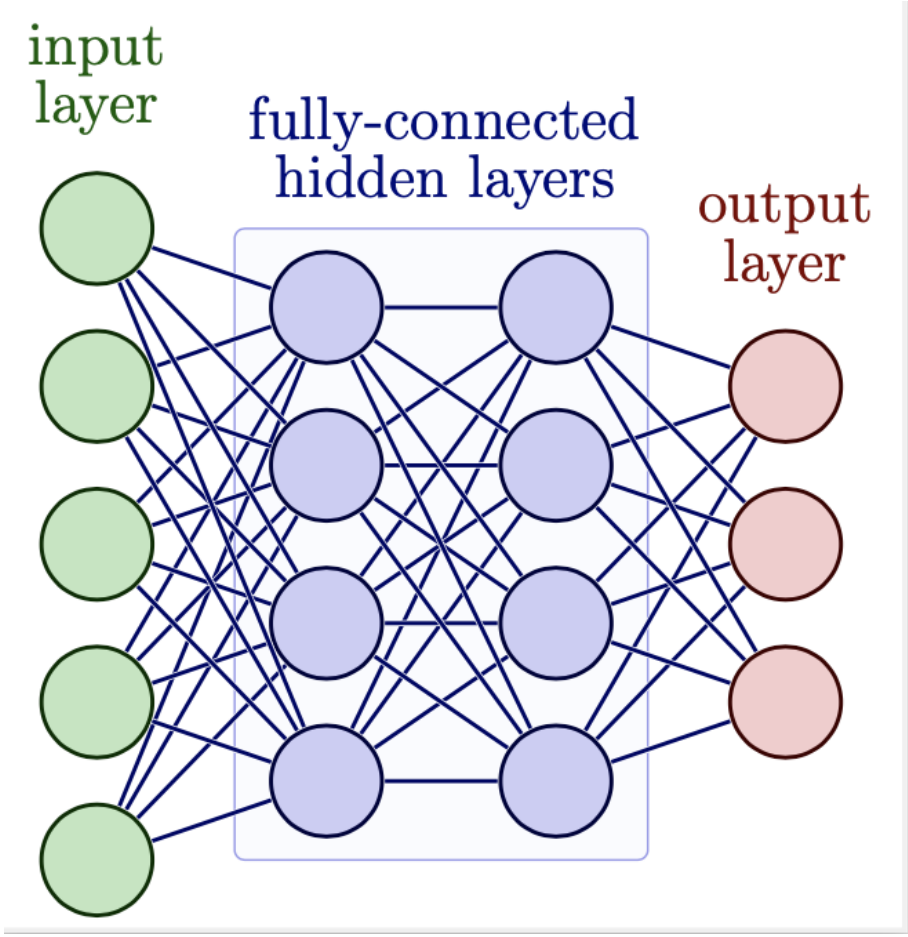
$$\text{prior} : \mathcal{GP}(m,k) \quad \Longrightarrow \quad \text{posterior} : \mathcal{GP}(\tilde{m}, \tilde{k})$$

NNPDF-like analysis

- generate $N_{\rm rep}$ replicas – NN initialized from random distributions
- NNs at initialization provide the prior, $f_i = N(x_i;\theta)$
- train the NNs using data
- NNs after training provide the posterior



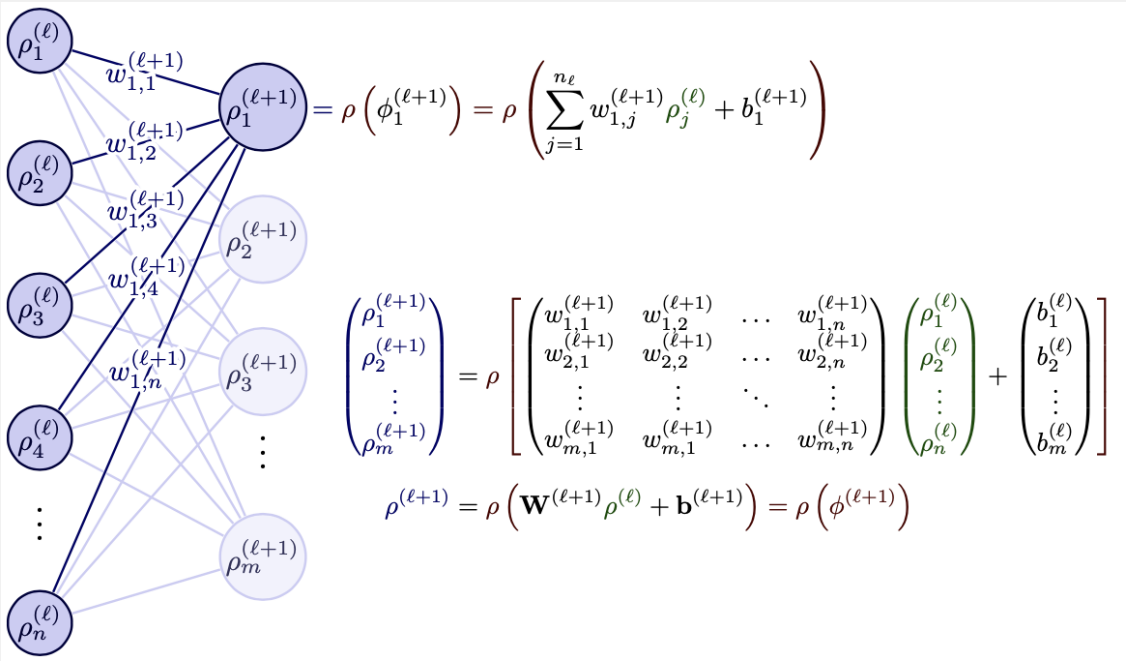
Neural Networks



MLP architecture

layers: $\ell = 1, \ldots, L$
neurons: $i = 1, \ldots, n_\ell$
weights $w_{ij}^{(\ell)}$, biases $b_i^{(\ell)}$
data: $(x_\alpha, y_\alpha), \alpha \in \mathcal{D}$

pre-activation functions



$$\boxed{\phi_{i\alpha}^{(\ell+1)} = \sum_{j=1}^{n_\ell} w_{ij}^{(\ell+1)} \rho_{j\alpha}^{(\ell)} + b_i^{(\ell+1)}}$$

statistical ensembles of NNs

initialize Gaussian weights and biases using Gaussians

$$\langle b_i^{(\ell)} \rangle = 0\,, \quad \langle b_{i_1}^{(\ell)} b_{i_2}^{(\ell)} \rangle = \delta_{i_1 i_2} C_b^{(\ell)}$$
$$\langle w_{ij}^{(\ell)} \rangle = 0\,, \quad \langle w_{i_1 j_1}^{(\ell)} w_{i_2 j_2}^{(\ell)} \rangle = \delta_{i_1 i_2} \delta_{j_1 j_2} \frac{C_w^{(\ell)}}{n_{\ell-1}}$$

parameters/functions duality

$$p(\phi^{(\ell)}|\mathcal{D}) = \int [dw\, p(w)]\, [db\, p(b)] \prod_{i,\alpha} \delta \left(\phi_{i\alpha}^{(\ell)} - \sum_j w_{ij}^{(\ell)} \rho \left(\phi_{j\alpha}^{(\ell-1)} \right) - b_i^{(\ell)} \right)$$

computing the integral

$$p(\phi^{(\ell+1)}|\mathcal{D}) = \int d\phi^{(\ell)}\, p(\phi^{(\ell+1)}|\phi^{(\ell)})\, p(\phi^{(\ell)}|\mathcal{D})$$

$$\begin{aligned} p(\phi^{(\ell+1)}|\phi^{(\ell)}) &= \int \left[dw^{(\ell+1)}\, p(w^{(\ell+1)})\right] \left[db^{(\ell+1)}\, p(b^{(\ell+1)})\right] \\ &\qquad \times \prod_{i,\alpha} \delta\big(\phi^{(\ell+1)}_{i\alpha} - \sum_j w^{(\ell+1)}_{ij} \rho\left(\phi^{(\ell)}_{j\alpha}\right) - b^{(\ell+1)}_i\big) \\ &= \frac{1}{|2\pi \widehat{G}^{(\ell+1)}|^{n_\ell/2}}\, \exp\left[-\frac{1}{2}\widehat{G}^{(\ell+1)}_{\alpha_1\alpha_2}\vec{\phi}^{(\ell+1)}_{\alpha_1} \cdot \vec{\phi}^{(\ell+1)}_{\alpha_2}\right] \end{aligned}$$

$$\langle \phi^{(\ell+1)}_{i_1\alpha_1} \ldots \phi^{(\ell+1)}_{i_{2k}\alpha_{2k}} \rangle = \sum_{\text{pairs}} \delta_{i_{P_1} i_{P_2}} \ldots \left\langle \left(\widehat{G}^{(\ell+1)}\right)^{-1}_{\alpha_{P_1}\alpha_{P_2}} \ldots \right\rangle$$

propagators and all that

$$\widehat{G}^{(\ell+1)}_{\alpha_1\alpha_2} = C_b^{(\ell+1)} + \frac{C_w^{(\ell+1)}}{n_\ell}\vec{\rho}^{(\ell)}_{\alpha_1}\cdot\vec{\rho}^{(\ell)}_{\alpha_2}$$

fluctuations of $\widehat{G}$

$$\widehat{\Delta G}^{(\ell+1)}_{\alpha_1\alpha_2} = \widehat{G}^{(\ell+1)}_{\alpha_1\alpha_2} - \langle\widehat{G}^{(\ell+1)}_{\alpha_1\alpha_2}\rangle$$

$$\langle\widehat{\Delta G}^{(\ell+1)}_{\alpha_1\alpha_2}\widehat{\Delta G}^{(\ell+1)}_{\alpha_3\alpha_4}\rangle = \frac{1}{n_\ell}V^{(\ell+1)}_{\alpha_1\alpha_2,\alpha_3\alpha_4}$$

correlators

$$\langle\phi^{(\ell+1)}_{i_1\alpha_1}\phi^{(\ell+1)}_{i_2\alpha_2}\rangle = \delta_{i_1i_2}\langle\widehat{G}^{(\ell+1)}_{\alpha_1\alpha_2}\rangle = \delta_{i_1i_2}G^{(\ell+1)}_{\alpha_1\alpha_2}$$

$$\langle\phi^{(\ell+1)}_{i_1\alpha_1}\phi^{(\ell+1)}_{i_2\alpha_2}\phi^{(\ell+1)}_{i_3\alpha_3}\phi^{(\ell+1)}_{i_4\alpha_4}\rangle_c = \frac{1}{n_\ell}\left[\delta_{i_1i_2}\delta_{i_3i_4}V^{(\ell+1)}_{\alpha_1\alpha_2,\alpha_3\alpha_4} + \ldots\right]$$

$1/n$ expansion

correlators can be expanded in $1/n$

$$G^{(\ell)}_{\alpha_1\alpha_2} = K^{(\ell)}_{\alpha_1\alpha_2} + \frac{1}{n_{\ell-1}} G^{\{1\}(\ell)}_{\alpha_1\alpha_2} + \frac{1}{n^2_{\ell-1}} G^{\{2\}(\ell)}_{\alpha_1\alpha_2} + O(\frac{1}{n^3_{\ell-1}})$$

$$V^{(\ell)}_{\alpha_1\alpha_2,\alpha_3\alpha_4} = V^{\{0\}(\ell)}_{\alpha_1\alpha_2,\alpha_3\alpha_4} + \frac{1}{n_{\ell-1}} V^{\{1\}(\ell)}_{\alpha_1\alpha_2,\alpha_3\alpha_4} + O(\frac{1}{n^2_{\ell-1}})$$

therefore

$$\langle \phi^{(\ell+1)}_{i_1\alpha_1} \phi^{(\ell+1)}_{i_2\alpha_2} \rangle = \delta_{i_1 i_2} K^{(\ell+1)}_{\alpha_1\alpha_2} + O(1/n)$$

$$\langle \phi^{(\ell+1)}_{i_1\alpha_1} \phi^{(\ell+1)}_{i_2\alpha_2} \phi^{(\ell+1)}_{i_3\alpha_3} \phi^{(\ell+1)}_{i_4\alpha_4} \rangle_c = \frac{1}{n_\ell} \left[\delta_{i_1 i_2} \delta_{i_3 i_4} V^{\{0\}(\ell+1)}_{\alpha_1\alpha_2,\alpha_3\alpha_4} + \ldots \right] + O(1/n^2)$$

EFT

probability described by an effective action

$$p(\phi|\mathcal{D}) = \frac{e^{-S(\phi)}}{Z}$$
$$S(\phi) = \frac{1}{2}\gamma^{(2)}_{\alpha_1\alpha_2}\,\vec{\phi}_{\alpha_1}\cdot\vec{\phi}_{\alpha_2} + \frac{1}{8}\gamma^{(4)}_{\alpha_1\alpha_2,\alpha_3\alpha_4}\,\vec{\phi}_{\alpha_1}\cdot\vec{\phi}_{\alpha_2}\,\vec{\phi}_{\alpha_3}\cdot\vec{\phi}_{\alpha_4} + \dots$$

couplings fixed by matching correlators in $1/n$ expansion

$$\begin{aligned}\langle\phi^{(\ell+1)}_{i_1\alpha_1}\phi^{(\ell+1)}_{i_2\alpha_2}\rangle &= \delta_{i_1i_2}K^{(\ell+1)}_{\alpha_1\alpha_2} + O(1/n)\\ &= \delta_{i_1i_2}\left(\gamma^{(2,(\ell+1))}\right)^{-1}_{\alpha_1\alpha_2} + O\left(\gamma^{(4,(\ell+1))}\right)\end{aligned}$$

$$\begin{aligned}\langle\phi^{(\ell+1)}_{i_1\alpha_1}\phi^{(\ell+1)}_{i_2\alpha_2}\phi^{(\ell+1)}_{i_3\alpha_3}\phi^{(\ell+1)}_{i_4\alpha_4}\rangle_c &= \delta_{i_1i_2}\delta_{i_3i_4}\frac{1}{n_\ell}V^{\{0\}(\ell+1)}_{\alpha_1\alpha_2,\alpha_3\alpha_4} + \dots\\ &= \delta_{i_1i_2}\delta_{i_3i_4}\left[G^{(\ell+1)}_{\alpha_1\beta_1}G^{(\ell+1)}_{\alpha_2\beta_2}G^{(\ell+1)}_{\alpha_3\beta_3}G^{(\ell+1)}_{\alpha_4\beta_4}\right]\gamma^{(4,(\ell+1))}_{\beta_1\beta_2,\beta_3\beta_4} + \dots\end{aligned}$$

recursion relations

two-pt function at leading order

$$K^{(\ell+1)}_{\alpha_1\alpha_2} = C_b^{(\ell+1)} + C_w^{(\ell+1)} \frac{1}{n_\ell} \langle \vec{\rho}^{(\ell)}_{\alpha_1} \cdot \vec{\rho}^{(\ell)}_{\alpha_2} \rangle \Big|_{O(1)}$$

$$= C_b^{(\ell+1)} + C_w^{(\ell+1)} \frac{1}{n_\ell} \langle \vec{\rho}^{(\ell)}_{\alpha_1} \cdot \vec{\rho}^{(\ell)}_{\alpha_2} \rangle_{K^{(\ell)}}$$

solving the recursion for $K_{\alpha\alpha}$

$$g(K) = \int d\phi \, \frac{1}{\sqrt{2\pi K}} \exp\big(-\frac{1}{2K}\phi^2\big) \rho(\phi)^2$$

fixed point: $K^* = C_b + C_w g(K^*)$

$$K_{\alpha\alpha} = K^* + \Delta K_{\alpha\alpha} \Longrightarrow \Delta K^{(\ell+1)}_{\alpha\alpha} = \chi_{\|}(K^*) \Delta K^{(\ell)}_{\alpha\alpha}$$

RG-style evolution as we go deep in the Neural Network

training

gradient descent

$$\frac{d}{dt}\theta_\mu(t) = -\lambda_{\mu\nu}\frac{\partial}{\partial\theta_\mu}\mathcal{L}_A$$

evolution of $O(\phi)$

$$\begin{aligned}\frac{d}{dt}O(t) &= -\frac{\partial O}{\partial\phi_{i\delta}}\frac{\partial\phi_{i\delta}}{\partial\theta_\mu}\lambda_{\mu\nu}\frac{\partial\mathcal{L}_A}{\partial\phi_{j\alpha}}\frac{\partial\phi_{j\alpha}}{\partial\theta_\nu} \\ &= -\left.\frac{\partial O}{\partial\phi_{i\delta}}\right|_{\phi(t)} H_{i\delta,j\alpha}(t)\,\varepsilon_{j\alpha}(t)\end{aligned}$$

in particular

$$\boxed{\frac{d}{dt}\phi_{i\delta}(t) = -H_{i\delta,j\alpha}(t)\varepsilon_{j\alpha}(t)}$$

EFT again

introduce an auxiliary field $L_{i\alpha}(t)$

$$p(\varphi, L|\mathcal{D}) = \frac{1}{Z}\exp\Big(-S(\phi) - \int dt\, L_{i\alpha}(t)\big(\frac{d}{dt}\varphi_{i\delta}(t) + H_{i\delta,j\alpha}(t)\varepsilon_{j\alpha}(t)\big)\Big)$$

with $\varphi_{i\alpha}(0) = \phi_{i\alpha}$

use this theory to compute correlators during training

$\langle L_{i_1\alpha_1}(t)\varphi_{i_2\alpha_2}(t')\rangle$: propagator

interactions depend on the loss

exact solution

quadratic loss

$$\mathcal{L}_A = \sum_{\alpha \in A} \big(y_\alpha - \phi_\alpha\big)^2$$

leading order in $1/n$

$$\frac{d}{dt} H = 0 \quad \Longrightarrow \quad \frac{d}{dt}\phi_{i\delta}(t) = 2H_{i\delta;j\alpha}\big(y_{j\alpha} - \phi_{j\alpha}\big)$$

discrete integration

$$\begin{aligned} \phi(n+1) &= \phi(n) - \eta H(\phi(n) - y) \\ &\Longrightarrow \phi(n) = \phi - Ha(n) \end{aligned}$$

$$a(n) = \eta \frac{1 - (1 - \eta \hat{H})^n}{1 - (1 - \eta \hat{H})} \big(\phi - y\big)$$

conclusions

- EFT description of NN at initialization
- $1/n$ expansion is important to have a hierarchy of couplings
- training is a flow in functions space
- NTK dictates the properties of the flow
- understand $1/n$ corrections to GP
- maybe useful for trivializing maps?
- work in progress