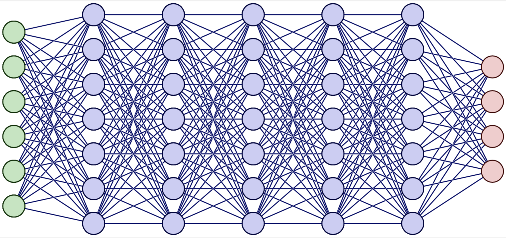
Field Theories and Machine Learning

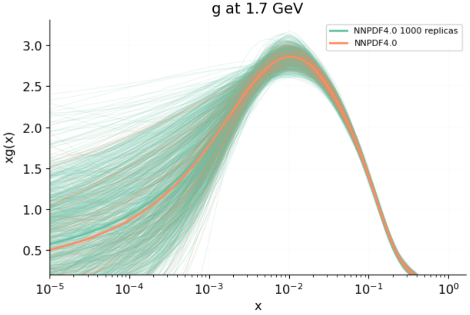


L Del Debbio

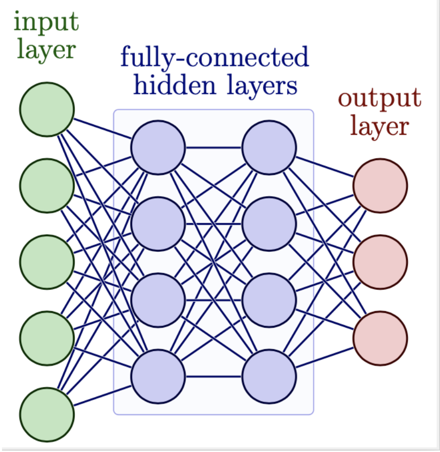
Higgs Centre for Theoretical Physics
University of Edinburgh

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



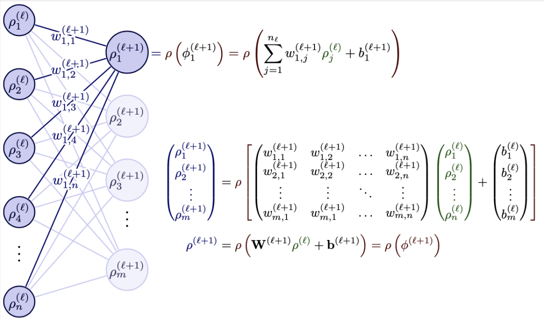
Neural Networks



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 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

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

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

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

covariance

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

fluctuations of $\widehat{G}$

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

$$\begin{aligned}
\langle \widehat{\Delta G}^{(\ell+1)}_{\alpha_1\alpha_2} \widehat{\Delta G}^{(\ell+1)}_{\alpha_3\alpha_4} \rangle &= \langle \widehat{G}^{(\ell+1)}_{\alpha_1\alpha_2} \widehat{G}^{(\ell+1)}_{\alpha_3\alpha_4} \rangle - \langle \widehat{G}^{(\ell+1)}_{\alpha_1\alpha_2} \rangle \langle \widehat{G}^{(\ell+1)}_{\alpha_3\alpha_4} \rangle \\
&= \frac{C_w^2}{n_\ell^2} \Big[\langle \vec{\rho}^{(\ell+1)}_{\alpha_1} \cdot \vec{\rho}^{(\ell+1)}_{\alpha_2} \vec{\rho}^{(\ell+1)}_{\alpha_3} \cdot \vec{\rho}^{(\ell+1)}_{\alpha_4} \rangle - \\
&\qquad - \langle \vec{\rho}^{(\ell+1)}_{\alpha_1} \cdot \vec{\rho}^{(\ell+1)}_{\alpha_2} \rangle \langle \vec{\rho}^{(\ell+1)}_{\alpha_3} \cdot \vec{\rho}^{(\ell+1)}_{\alpha_4} \rangle \Big] \\
&= \frac{1}{n_\ell} V^{(\ell+1)}_{\alpha_1\alpha_2,\alpha_3\alpha_4}
\end{aligned}$$

propagators and all that

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

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

$$\langle \phi^{(\ell+1)}_{i_1\alpha_1} \phi^{(\ell+1)}_{i_2\alpha_2} \phi^{(\ell+1)}_{i_3\alpha_3} \phi^{(\ell+1)}_{i_4\alpha_4} \phi^{(\ell+1)}_{i_5\alpha_5} \phi^{(\ell+1)}_{i_6\alpha_6} \rangle_c = \frac{1}{n_\ell^2} \times \dots$$

$1/n$ 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

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

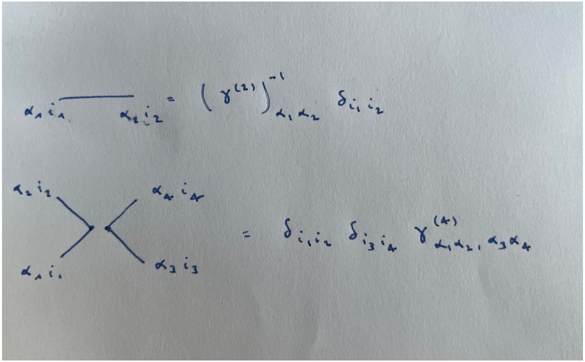
$$\langle \phi^{(\ell+1)}_{i_1\alpha_1} \phi^{(\ell+1)}_{i_2\alpha_2} \phi^{(\ell+1)}_{i_3\alpha_3} \phi^{(\ell+1)}_{i_4\alpha_4} \rangle_c = \frac{1}{n_\ell} \left[\delta_{i_1 i_2} \delta_{i_3 i_4} V^{\{0\}(\ell+1)}_{\alpha_1\alpha_2,\alpha_3\alpha_4} + \dots \right] + O(1/n^2)$$

EFT

probability described by an effective action

$$p(\phi|\mathcal{D}) = \frac{e^{-S(\phi)}}{Z}$$

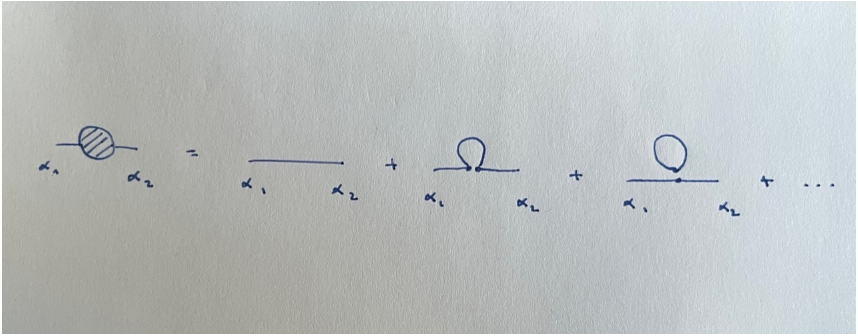
$$S(\phi) = \frac{1}{2}\gamma^{(2)}_{\alpha_1\alpha_2}\,\vec{\phi}_{\alpha_1}\cdot\vec{\phi}_{\alpha_2} + \frac{1}{8}\gamma^{(4)}_{\alpha_1\alpha_2,\alpha_3\alpha_4}\,\vec{\phi}_{\alpha_1}\cdot\vec{\phi}_{\alpha_2}\,\vec{\phi}_{\alpha_3}\cdot\vec{\phi}_{\alpha_4} + \dots$$



Two-point Function

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

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

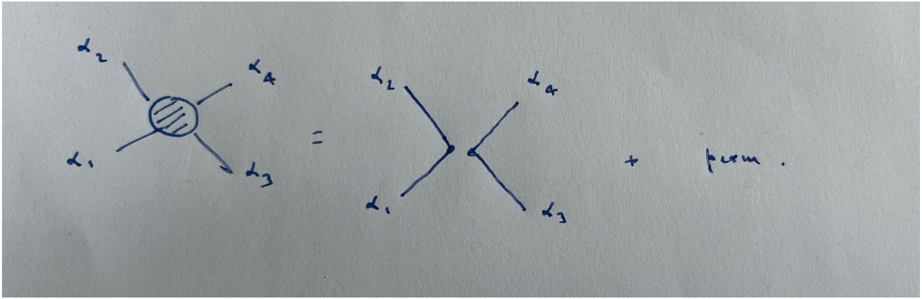


Four-point Function

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

$$\langle \phi^{(\ell+1)}_{i_1\alpha_1} \phi^{(\ell+1)}_{i_2\alpha_2} \phi^{(\ell+1)}_{i_3\alpha_3} \phi^{(\ell+1)}_{i_4\alpha_4} \rangle_c = \delta_{i_1 i_2} \delta_{i_3 i_4} \frac{1}{n_\ell} V^{\{0\}(\ell+1)}_{\alpha_1\alpha_2,\alpha_3\alpha_4} + \dots$$

$$= \delta_{i_1 i_2} \delta_{i_3 i_4} \left[G^{(\ell+1)}_{\alpha_1\beta_1} G^{(\ell+1)}_{\alpha_2\beta_2} G^{(\ell+1)}_{\alpha_3\beta_3} G^{(\ell+1)}_{\alpha_4\beta_4} \right] \gamma^{(4,(\ell+1))}_{\beta_1\beta_2,\beta_3\beta_4} + \dots$$



going deep – recursion relations

two-pt function at leading order

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

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

$$\frac{1}{n_\ell} \langle \vec{\rho}^{(\ell)}_{\alpha_1} \cdot \vec{\rho}^{(\ell)}_{\alpha_2} \rangle_{K^{(\ell)}} = \int \prod_\alpha d\phi_\alpha \, \frac{e^{-\frac{1}{2}\left(K^{(\ell)}\right)^{-1}_{\beta_1\beta_2} \phi_{\beta_1}\phi_{\beta_2}}}{\left|2\pi K^{(\ell)}\right|^{1/2}} \, \rho(\phi_{\alpha_1})\rho(\phi_{\alpha_2})$$

evaluating the integral

if $\alpha_1 = \alpha_2$, only one input

$$\int \prod_\alpha d\phi_\alpha \, \frac{e^{-\frac{1}{2}\left(K^{(\ell)}\right)^{-1}_{\beta_1\beta_2}\phi_{\beta_1}\phi_{\beta_2}}}{\left|2\pi K^{(\ell)}\right|^{1/2}} \, \rho(\phi_{\alpha_1})\rho(\phi_{\alpha_1}) =$$

$$= \int d\phi_{\alpha_1} \, \frac{e^{-\frac{1}{2}\left(K^{(\ell)}_{\alpha_1\alpha_1}\right)^{-1}\phi^2_{\alpha_1}}}{\left(2\pi K^{(\ell)}_{\alpha_1\alpha_1}\right)^{1/2}} \, \rho(\phi_{\alpha_1})^2 =$$

$$= g(K^{(\ell)}_{\alpha_1\alpha_1})$$

solving the recursion

$$K_{\alpha\alpha}^{(\ell)} = \langle \frac{1}{n_\ell} \sum_{i=0}^{n_\ell} \left(\phi_{i,\alpha}^{(\ell)}\right)^2 \rangle$$

$K_{\alpha\alpha}$, introduce the auxiliary function

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

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

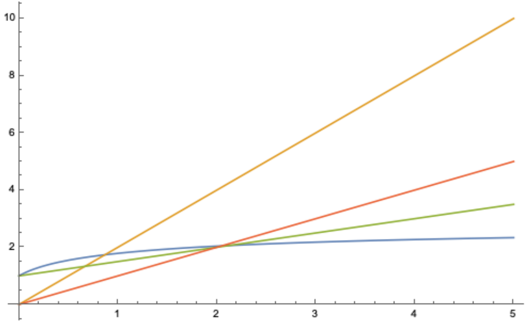
$$K_{\alpha\alpha} = K^* + \Delta K_{\alpha\alpha} \Longrightarrow \Delta K_{\alpha\alpha}^{(\ell+1)} = \chi_{\|}(K^*) \Delta K_{\alpha\alpha}^{(\ell)}$$

$$\chi_{\|}(K) = C_w g'(K)$$

criticality: $\chi_{\|}(K^*) = 1$

examples

graphical solution for $C_b = 1, C_w = 2$



more recursions

for the off-diagonal elements of K: $\alpha_1 \neq \alpha_2$

$$K^{(\ell+1)}_{\alpha_1\alpha_2} = C_b^{(\ell+1)} + C_w^{(\ell+1)} \times$$
$$\times \int d\phi_{\alpha_1} d\phi_{\alpha_2} e^{-\frac{1}{2}(\phi_{\alpha_1}\phi_{\alpha_2})\bar{K}^{-1}\binom{\phi_{\alpha_1}}{\phi_{\alpha_2}}} \rho(\phi_{\alpha_1})\rho(\phi_{\alpha_2})$$

$$= H(K^{(\ell)}_{\alpha_1\alpha_1}, K^{(\ell)}_{\alpha_2\alpha_2}, K^{(\ell)}_{\alpha_1\alpha_2})$$

$$\bar{K} = \begin{pmatrix} K_{\alpha_1\alpha_1} & K_{\alpha_1\alpha_2} \\ K_{\alpha_2\alpha_1} & K_{\alpha_2\alpha_2} \end{pmatrix}$$

$\hookrightarrow$ coupled equations

RG-style evolution equation to go deep into the network

RG interpretation

- distribution in each layer at initialization is described by an EFT
- for $n \to \infty$, NN defines a GP
- EFT couplings obey RG relations as we go *deep*
- taken into account in the choice of the architecture/initialization parameters C_w and C_b

Bayesian Learning with NN

Divide the data into a training set and the rest

$$\mathcal{D} = \mathcal{A} \cup \mathcal{B}$$

Bayes:

$$p(\theta|\mathcal{A}, \mathcal{H}) = \frac{p(\mathcal{A}|\theta, \mathcal{H}) p(\theta|\mathcal{H})}{p(\mathcal{A}|\mathcal{H})}$$

likelihood:

$$p(A|\theta, \mathcal{H}) = \exp\left(-\mathcal{L}(\theta, \mathcal{A})\right)$$

estimators:

$$\theta^*_{\mathsf{MAP}} = \arg\max_\theta p(\theta|\mathcal{A}, \mathcal{H})$$

$$\theta^*_{\mathsf{MLE}} = \arg\max_\theta p(y_A|\theta, \mathcal{H})$$

Bayesian Learning at ∞ Width

$$p(\phi_B^{(L)}|y_a) = \frac{p(y_A, \phi_B^{(L)})}{p(y_A)}$$

leading order in $1/n$

$$p(y_A, \phi_B^{(L)}) \propto \exp\left(-\frac{1}{2}(y_A, \phi^{(L)})^T K^{-1} \begin{pmatrix} y_A \\ \phi_B^{(L)} \end{pmatrix}\right)$$

where

$$K^{-1} = \begin{pmatrix} K^{-1}_{\alpha\alpha} & K^{-1}_{\alpha\beta} \\ K^{-1}_{\beta\alpha} & K^{-1}_{\beta\beta} \end{pmatrix}, \quad K^{-1}_{\beta\beta} = K_{\beta\beta} - K_{\beta\alpha}(K_{\alpha\alpha})^{-1} K_{\alpha\beta}$$

$$p(\phi_B^{(L)}|y_a) \propto \exp\left(-\frac{1}{2}(\phi^{(L)} - m_\beta^\infty)^T K^{-1}_{\beta\beta}(\phi^{(L)} - m_\beta^\infty)\right)$$

$$m_\beta^\infty = K_{\beta\alpha}(K_{\alpha\alpha})^{-1} y_\alpha$$

Lack of Representation Learning

at infinite width

$$p(\phi^{(L)}_{\beta,1},\ldots,\phi^{(L)}_{\beta,n_L}|y_A) = p(\phi^{(L)}_{\beta,1}|y_A)\ldots p(\phi^{(L)}_{\beta,n_L}|y_A)$$

different neurons on the output layer are statistically independent

$$p(\phi_D^{(L-1)}|y_A) = \frac{p(y_A|\phi_D^{(L-1)})p(\phi_D^{(L-1)})}{p(y_A)}$$

$$p(y_A|\phi_D^{(L-1)}) \propto \exp\left(-\frac{1}{2}y^T(K_{\alpha\alpha})^{-1}y\right) = p(y_A)$$

$$p(\phi_D^{(L-1)}|y_A) = p(\phi_D^{(L-1)})$$

Bayesian Learning at finite width

consider a single datapoint, quartic interactions introduce correlations

$$p(\phi_1, \dots, \phi_m) \propto \exp\left(-\frac{\gamma^{(2)}}{2}\phi^T\phi + \frac{\gamma^{(4)}}{8}(\phi^T\phi)^2\right)$$

$$\frac{1}{\gamma^{(2)}} = G^{(\ell)} - \frac{m+2}{n_{\ell-1}}\frac{V^{(\ell)}}{G^{(\ell)}} + O\left(1/n^2\right)$$

$$\gamma^{(4)} = \frac{1}{n_{\ell-1}}\frac{V^{(\ell)}}{(G^{(\ell)})^4} + O\left(1/n^2\right)$$

conditional distribution for ϕ_2

$$p(\phi_2|\check{\phi}_1) \propto \exp\left(-\frac{\gamma^{(2)}}{2}\phi_2^2 + \frac{\gamma^{(4)}}{8}(\phi_2^4 + 2\phi_2^2\check{\phi}_1^2)\right)$$

$$\int d\phi_2\, p(\phi_2|\check{\phi}_1)\phi_2^2 = G^{(\ell)} + \frac{1}{2}\left(\check{\phi}_1^2 - G^{(\ell)}\right)\frac{V^{(\ell)}}{n_{\ell-1}(G^{(\ell)})^2} + O(1/n^2)$$

training and NTK

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

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

Forward Equation for the NTK

$$H^{(\ell)} = \lambda_{\mu\nu} \frac{\partial \phi^{(\ell)}}{\partial \theta_\mu} \frac{\partial \phi^{(\ell)}}{\partial \theta_\nu}$$

learning rate tensor

$$\lambda_{b^{(\ell)}_{i_1} b^{(\ell)}_{i_2}} = \delta_{i_1 i_2} \lambda^{(\ell)}_b \, , \quad \lambda_{W^{(\ell)}_{i_1 j_1} W^{(\ell)}_{i_2 j_2}} = \delta_{i_1 i_2} \delta_{j_1 j_2} \frac{\lambda^{(\ell)}_W}{n_{\ell-1}}$$

then

$$H^{(\ell)}_{i_1\alpha_1, i_2\alpha_2} = \sum_{\ell'=1}^{\ell} \left(\lambda_b^{(\ell')} \frac{\partial \phi^{(\ell')}_{i_1\alpha_1}}{\partial b^{(\ell')}_j} \frac{\partial \phi^{(\ell')}_{i_2\alpha_2}}{\partial b^{(\ell')}_j} + \frac{\lambda_w^{(\ell')}}{n_{\ell-1}} \frac{\partial \phi^{(\ell')}_{i_1\alpha_1}}{\partial W^{(\ell')}_{jk}} \frac{\partial \phi^{(\ell')}_{i_2\alpha_2}}{\partial W^{(\ell')}_{jk}} \right)$$

$$
\begin{aligned}
H^{(\ell+1)}_{i_1 i_2 \alpha_1 \alpha_2} &= \sum_{j=1}^{n_{\ell+1}} \left(\lambda_b^{(\ell+1)} \frac{\partial \phi^{(\ell+1)}_{i_1 \alpha_1}}{\partial b_j^{(\ell+1)}} \frac{\partial \phi^{(\ell+1)}_{i_2 \alpha_2}}{\partial b_j^{(\ell+1)}} + \frac{\lambda_w^{(\ell+1)}}{n_{\ell-1}} \frac{\partial \phi^{(\ell+1)}_{i_1 \alpha_1}}{\partial W_{jk}^{(\ell+1)}} \frac{\partial \phi^{(\ell+1)}_{i_2 \alpha_2}}{\partial W_{jk}^{(\ell+1)}} \right) + \\
&\quad + \sum_{j_1, j_2 = 1}^{n_\ell} \frac{\partial \phi^{(\ell+1)}_{i_1 \alpha_1}}{\partial \phi^{(\ell)}_{j_1 \alpha_1}} \frac{\partial \phi^{(\ell+1)}_{i_2 \alpha_2}}{\partial \phi^{(\ell)}_{j_2 \alpha_2}} H^{(\ell)}_{j_1 j_2 \alpha_1 \alpha_2} \\
&= \delta_{i_1 i_2} \left[\lambda_b^{(\ell+1)} + \lambda_W^{(\ell+1)} \left(\frac{1}{n_\ell} \sum_{j=1}^{n_\ell} \rho^{(\ell)}_{j \alpha_1} \rho^{(\ell)}_{j \alpha_2} \right) \right] + \\
&\quad + \sum_{j_1, j_2 = 1}^{n_\ell} W^{(\ell+1)}_{i_1 j_1} W^{(\ell+1)}_{i_2 j_2} \rho'^{(\ell)}_{j_1 \alpha_1} \rho'^{(\ell)}_{j_2 \alpha_2} H^{(\ell)}_{j_1 j_2 \alpha_1 \alpha_2}
\end{aligned}
$$

$1/n$ expansion

$$
H = \Theta + \frac{1}{n} H^{[1]} + \dots
$$

Gradient Training at Infinite Width

at the minimum of $\mathcal{L}(\theta, A)$

$$\phi^{(L)}(T) = \phi^{(L)} - \Theta^{(L)}_{\beta\alpha}(\Theta^{(L)}_{\alpha\alpha})^{-1}\left(\phi^{(L)} - y\right)$$

hence

$$m^{\infty} = \Theta^{(L)}_{\beta\alpha}(\Theta^{(L)}_{\alpha\alpha})^{-1} y$$

recall

$$\begin{aligned} K^{(L)}_{\delta_1\delta_2} &= C^{(L)}_b + C^{(L)}_W \langle \rho_{\delta_1}\rho_{\delta_2}\rangle_{K^{(L-1)}} \\ \Theta^{(L)}_{\delta_1\delta_2} &= \lambda^{(L)}_b + \lambda^{(L)}_W \langle \rho_{\delta_1}\rho_{\delta_2}\rangle_{K^{(L-1)}} + C^{(L)}_W \langle \rho'_{\delta_1}\rho'_{\delta_2}\rangle_{K^{(L-1)}} \Theta^{(L-1)}_{\delta_1\delta_2} \end{aligned}$$

$$\Longrightarrow \lambda^{(L)}_b = C^{(L)}_b\,, \quad \lambda^{(L)}_W = C^{(L)}_W$$

EFT again

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

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

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

EOM:

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

hence

$$\frac{d}{dt}\mathcal{L}_A = -\varepsilon_{i\delta} H_{i\delta,j\alpha}(t)\varepsilon_{j\alpha}(t)$$

use this theory to compute correlators during training

Summary

- distribution of $\phi_{i\alpha}$ can be described by EFT
- power counting scheme in $1/n$ to have a predictive framework
- NN at infinite width $\longrightarrow$ made contact with GP
- $1/n$ corrections to go beyond Gaussianity
- deep networks and RG equations
- training and NTK analysis