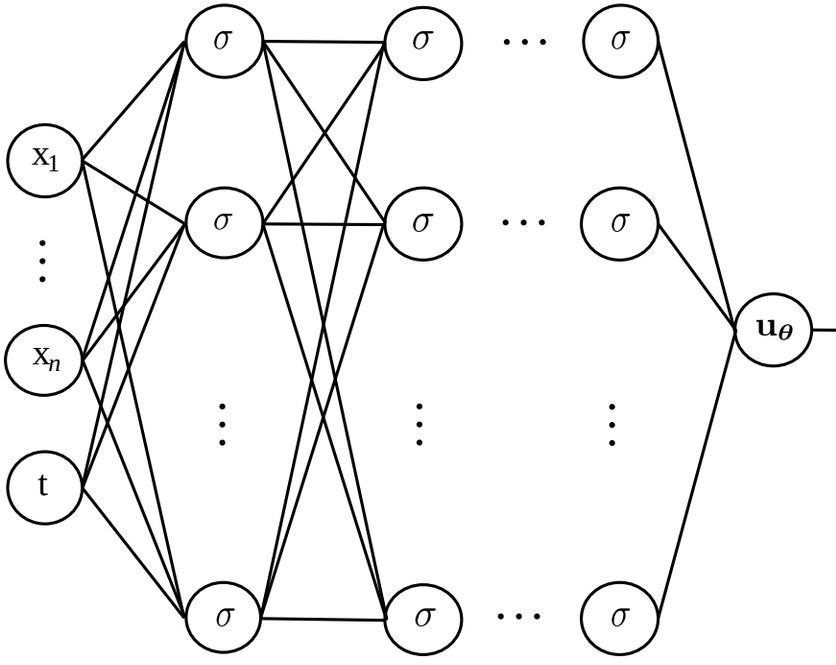
Pre-training phase



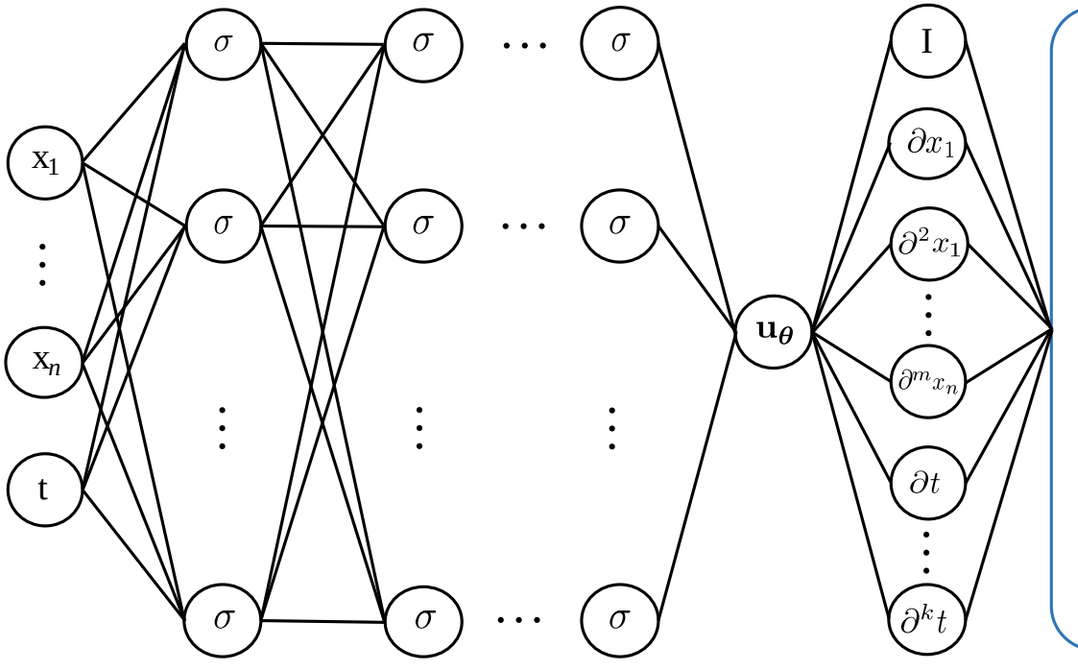
Data loss:

$$\mathcal{L}_d(\boldsymbol{\theta}) = \frac{1}{N_d}\sum_{k=1}^{N_d}\left(u_{\boldsymbol{\theta}}\left(\boldsymbol{x}_d^k, t_d^k\right) - u\left(\boldsymbol{x}_d^k, t_d^k\right)\right)^2$$

Minimize:

$$\boldsymbol{\theta}_d = \arg\min_{\boldsymbol{\theta}} \mathcal{L}_d(\boldsymbol{\theta})$$

Fine-tuning phase



PDE residual loss:

$$\mathcal{L}_r(\boldsymbol{\Theta}) = \frac{1}{N_r}\sum_{k=1}^{N_r}\left(\mathcal{F}\left[u_{\boldsymbol{\theta}}\left(\boldsymbol{x}_r^k, t_r^k\right);\boldsymbol{\gamma}\right] - s\left(\boldsymbol{x}_r^k, t_r^k\right)\right)^2$$

Initial condition loss:

$$\mathcal{L}_i(\boldsymbol{\theta}) = \frac{1}{N_i}\sum_{k=1}^{N_i}\left(\mathcal{I}\left[u_{\boldsymbol{\theta}}\left(\boldsymbol{x}_i^k, t_i^k\right)\right] - f\left(\boldsymbol{x}_i^k\right)\right)^2$$

Boundary condition loss:

$$\mathcal{L}_b(\boldsymbol{\theta}) = \frac{1}{N_b}\sum_{k=1}^{N_b}\left(\mathcal{B}\left[u_{\boldsymbol{\theta}}\left(\boldsymbol{x}_b^k, t_b^k\right)\right] - q\left(\boldsymbol{x}_b^k, t_b^k\right)\right)^2$$

Data loss:

$$\mathcal{L}_d(\boldsymbol{\theta}) = \frac{1}{N_d}\sum_{k=1}^{N_d}\left(u_{\boldsymbol{\theta}}\left(\boldsymbol{x}_d^k, t_d^k\right) - u\left(\boldsymbol{x}_d^k, t_d^k\right)\right)^2$$

Minimize:

$$\boldsymbol{\Theta}^\star = \arg\min_{\boldsymbol{\Theta}}\left[\lambda_r\mathcal{L}_r(\boldsymbol{\Theta}) + \lambda_i\mathcal{L}_i(\boldsymbol{\theta}) + \lambda_b\mathcal{L}_b(\boldsymbol{\theta}) + \lambda_d\mathcal{L}_d(\boldsymbol{\theta})\right]$$

$$\text{subject to } \boldsymbol{\theta}_0 = \boldsymbol{\theta}_d$$