hyper-parameter $\{\beta_t\}_{t=1\ldots\infty}$ ▷ exploration-exploitation trade-off values
hyper-parameter n_p ▷ random samples used to pre-train the model
hyper-parameter r ▷ Hessian approximation rank
hyper-parameter n_H ▷ samples used for training Hessian approximation
hyper-parameter n_I ▷ samples used for computing influence

procedure NNINF(f, $\mathcal{X}$, T, g_θ)
▶ Minimize f over $\mathcal{X}$ for T steps using the network g_θ.

$D \leftarrow \{(x, f(x)) : x \in \textsc{Sample}(\mathcal{X}, n_p)\}$ ▷ samples for pre-training
$|\theta| \leftarrow$ number of parameters in θ
$P \leftarrow \textsc{Matrix}(|\theta|, r)$ ▷ for low rank Hessian approximation

for $t \leftarrow 1 \ldots T$ **do**
$\textsc{TrainNetwork}(g_\theta, D)$
$P, \mathcal{I} \leftarrow \text{IHVP}(g_\theta, D, P)$
$x_t \leftarrow \arg\min_{x \in \mathcal{X}} \textsc{Acquisition}(x, g_\theta, \mathcal{I}, \beta_t)$
$D \leftarrow D \cup \{(x_t, f(x_t))\}$
end for

return $\arg\min_{(x,y)\in D} y$
end procedure

procedure IHVP(g_θ, D, P)
▶ Compute $H_\theta^{-1} \nabla_\theta L(z, \theta)$ for $z \in D$.

$\pi_P \leftarrow \textsc{FullyConnectedNetwork}\big(P, P^T\big)$
$L_H \leftarrow \{\nabla_\theta L(z, \theta) : z \in \textsc{Sample}(D, n_H)\}$
$J_\theta \leftarrow (1/n_H) \sum_{v \in L_H} v$
$\nu_J \leftarrow \nabla_\theta J_\theta$
$D_H \leftarrow \{\big(v, \nabla_\theta v^T \nu_J\big) : v \in L_H\}$

$\textsc{TrainNetwork}(\pi_P, D_H)$
$U, \Sigma, V \leftarrow \text{SVD}(P)$
$W \leftarrow U {\Sigma^\dagger}^2$

$\mathcal{I} \leftarrow \{W U^T v : v \in \textsc{Sample}(L_H, n_I)\}$
return $P, \mathcal{I}$
end procedure

procedure $\textsc{Acquisition}(x, g_\theta, \mathcal{I}, \beta)$
▶ compute the acquisition function at x

$\mu \leftarrow g_\theta(x)$
$\nu_\mu \leftarrow \nabla_\theta \mu$
$\sigma \leftarrow \sqrt{\frac{1}{n_I} \sum_{\iota \in \mathcal{I}} \left(\mathbf{1}^T \nu_\mu \iota\right)^2}$
return $\mu - \beta^{1/2} \sigma$
end procedure