update parameters θ

h_1 → MLP → LSTM → MLP → $Q_1^{\mathrm{rob}}(h_1, \cdot; \theta)$

h_N → MLP → LSTM → MLP → $Q_N^{\mathrm{rob}}(h_N, \cdot; \theta)$

$(Q_i^{\mathrm{rob}})_{i \in [N]}$ satisfy **DrIGM** w.r.t. $Q_{\mathrm{tot}}^{\mathcal{P}}$

ϵ-greedy

$\mathbf{a}$

P^0

$\mathbf{h}'$ → $\mathbf{a}'$

$Q_{\mathrm{tot}}^{\mathcal{P}}(\mathbf{h}, \mathbf{a}; \theta)$

$Q_{\mathrm{tot}}^{\mathcal{P}}(\mathbf{h}', \mathbf{a}'; \theta^-)$

target network

robust loss L_{TD}

robustness parameter ρ

$a_i := \underset{a \in \mathcal{A}_i}{\arg\max}\, Q_i^{\mathrm{rob}}(h_i, a; \theta)$

$a_i' := \underset{a \in \mathcal{A}_i}{\arg\max}\, Q_i^{\mathrm{rob}}(h_i', a; \theta^-)$