

Target Propagation (R. & Harchaoui 2021)

Idea Le Cun et al. (1988), Lee et al. (2015), Meulemans et al. (2020)

1. Back-propagate “targets” through inverted layer

$$\underbrace{h_{t+1} = f_{\theta}(h_t)}_{\text{forward pass}} \rightarrow \underbrace{v_{t-1} = f_{\theta}^{-1}(v_t)}_{\text{backward pass}}$$

starting from $v_L = h_L - \nabla_{h_L} \ell(y, h_L)$, $h_0 = x$

2. Try aligning layers with targets by GD step

$$\theta \leftarrow \theta - \sum_{t=1}^{L-1} \frac{\partial \|v_{t+1} - f_{\theta}(h_t)\|_2^2}{\partial \theta}$$

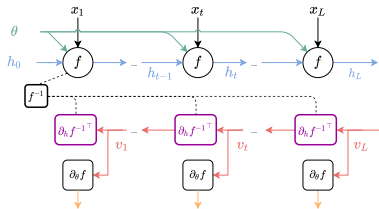
Inverse layers

Usual: learned auto-encoder

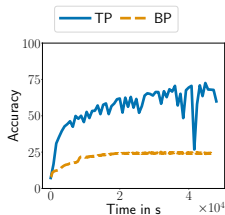
Proposed: variational regularized formulation

Formulation used in practice

$$v_{t-1} = h_{t-1} + \nabla f_{\theta}^{-1}(h_t)^{\top} (v_t - h_t)$$



Target Propagation for RNN



MNIST pixel by pixel
TP: Target Propagation
BP: Gradient Back-Propagation