Neural Uncertainty

Max Cohen

September 14, 2020

Chapter 1

Model definition

We define a L layer RNN followed by a fully connected layer. At time step t,

$$\begin{cases} x_{t+1} = tanh(W_x h^L_{t+1} + b_x) + \epsilon_t \\ h^l_{t+1} = tanh(W^l_{hy} h^{l-1}_{t+1} + W^l_{hh} h^l_t + b^l_h) + \eta_t \quad \forall 1 \leq l \leq L \end{cases}$$

and $h^0_t \equiv y_t$.

Let's consider the weights of the last RNN layer as $\Theta^L \equiv (W^L_{hy}, W^L_{hh}, b^L_h)$, we have $h^L_{t+1} = f_{\Theta^L}(h^{L-1}_{t+1}, h^L_t) + \eta_t$.

$$Q(\hat{\Theta}^L_p, \Theta^L) = \mathbb{E}_{\hat{\Theta}^L_p}\left[log\ p_{\Theta^L}(X_{1:T}, h_{1:T}, y_{1:T}|X_{1:T})\right] \tag{1.1}$$

$$\simeq \frac{1}{M}\sum_{i=1}^{M} log\ p_{\Theta^L}(X_{1:T}, h_{1:T}, y_{1:T}) \tag{1.2}$$