Neural Uncertainty

Max Cohen

September 17, 2020

Chapter 1

Model definition

1.1 Notations

We consider the prediction task of a hidden state $(x^1, \cdots x^T)$ given a set of observations $(y^1, \cdots y^T)$.

1.2 Model

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

$$\begin{cases} x_{t+1} = \tanh(W_x h_{t+1}^L + b_x) \\ h_{t+1}^l = \tanh(W_{hy}^l h_{t+1}^{l-1} + W_{hh}^l h_t^l + b_h^l) \quad \forall 1 \leq l \leq L \end{cases}$$

with $h_t^0 \equiv y_t \; \forall t$ and $h_0^l \equiv 0 \; \forall 1 \leq l \leq L$.

Let's consider the weights of the last RNN and fully connected layers as $\Theta \equiv (W_{hy}^L, W_{hh}^L, b_h^L, W_x, b_x)$. We can define a new matrix h_t at each time step corresponding to the concatenation of all RNN layers: $h_t \equiv (h_t^1 \cdots h_t^L)$. We also introduce two sequences of random noises as i.i.d real valued random variables ϵ and η. We can now write our model in terms of two functions f and g as:

$$\begin{cases} x_{t+1} = f_\Theta(h_t) + \epsilon_t & \text{observation model} \\ h_{t+1} = g_\Theta(h_t, y_t) + \eta_t & \text{state model} \end{cases} \tag{1.1}$$

$$
\begin{aligned}
Q(\hat{\Theta}_p, \Theta) &= \mathbb{E}_{\hat{\Theta}_p} \left[\log\ p_\Theta(X_{1:T}, h_{1:T}, y_{1:T}) | X_{1:T} \right] \\
&\simeq \frac{1}{M} \sum_{i=1}^{M} log\ p_\Theta(X_{1:T}, h_{1:T}, y_{1:T})
\end{aligned}
$$