

Chapter 1

Model definition

1.1 Notations

We consider the prediction task of a set of observations (y^1, \dots, y^T) given a set of input (u^1, \dots, u^T) .

1.2 Model

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

$$\begin{cases} y_{t+1} = \tanh(W_y x_{t+1}^L + b_y) \\ x_{t+1}^l = \tanh(W_{xx}^l x_t^l + W_{xu}^l x_{t+1}^{l-1} + b_x^l) \quad \forall 1 \leq l \leq L \end{cases}$$

with $x_t^0 \equiv u_t \forall t$ and $x_0^l \equiv 0 \forall 1 \leq l \leq L$.

Let's consider the weights of the last RNN and fully connected layers as $\Theta \equiv (W_{xx}^L, W_{xu}^L, b_x^L, W_y, b_y)$. We can define a new matrix y_t at each time step corresponding to the concatenation of all RNN layers: $x_t \equiv (x_t^1 \dots x_t^L)$. 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:

$$\begin{cases} y_{t+1} = f_{\Theta}(x_{t+1}) + \epsilon_{t+1} & \text{observation model} \\ x_{t+1} = g_{\Theta}(x_t, u_{t+1}) + \eta_{t+1} & \text{state model} \end{cases} \quad (1.1)$$

$$Q(\hat{\Theta}_p, \Theta) = \mathbb{E}_{\hat{\Theta}_p} [\log p_{\Theta}(X_{1:T}, y_{1:T}, u_{1:T}) | y_{1:T}] \quad (1.2)$$

Let's develop the interior of the expectation:

$$\begin{aligned}
p_{\Theta}(X_{1:T}, y_{1:T}, u_{1:T}) &= \frac{1}{T} \log \left(\prod_{k=1}^T p_{\Theta}(x_k | x_{k-1}, u_k) p_{\Theta}(y_k | x_k) \right) \\
&= \frac{1}{T} \sum_{k=1}^T \log p_{\Theta}(x_k | x_{k-1}, u_k) + \log p_{\Theta}(y_k | x_k) \\
&= \frac{1}{T} \sum_{k=1}^T \log \left(\det(2\pi\Sigma_x)^{-1/2} \exp\left(-\frac{1}{2}(x_k - g_{\Theta}(x_{k-1}, u_k))^T \Sigma_x^{-1} (x_k - g_{\Theta}(x_{k-1}, u_k))\right) \right) \\
&\quad + \log \left(\det(2\pi\Sigma_y)^{-1/2} \exp\left(-\frac{1}{2}(y_k - f_{\Theta}(x_k))^T \Sigma_y^{-1} (y_k - f_{\Theta}(x_k))\right) \right) \\
&= -\frac{1}{2} \log |\Sigma_x| - \frac{1}{2} \log |\Sigma_y| \\
&\quad - \frac{1}{2T} \sum_{k=1}^T (x_k - g_{\Theta}(x_{k-1}, u_k))^T \Sigma_x^{-1} (x_k - g_{\Theta}(x_{k-1}, u_k)) \\
&\quad - \frac{1}{2T} \sum_{k=1}^T (y_k - f_{\Theta}(x_k))^T \Sigma_y^{-1} (y_k - f_{\Theta}(x_k))
\end{aligned}$$

In order to find the minimal value of Σ_x , we can search for the zeros of the derivate of the convexe function $\Sigma_x \mapsto p_{\Theta}(X_{1:T}, y_{1:T}, u_{1:T})$.

$$\begin{aligned}
\frac{\partial p_{\Theta}(X_{1:T}, y_{1:T}, u_{1:T})}{\partial \Sigma_x^{-1}} &= \frac{1}{2} \Sigma_x - \frac{1}{2T} \sum_{k=1}^T 2 \Sigma_x^{-1} (x_k - f_{\Theta}(x_{k-1}, u_k)) \\
&= \frac{1}{2} \Sigma_x - \frac{\Sigma_x^{-1}}{T} \sum_{k=1}^T x_k - f_{\Theta}(x_{k-1}, u_k)
\end{aligned}$$

$$\frac{\partial p_{\Theta}(X_{1:T}, y_{1:T}, u_{1:T})}{\partial \Sigma_x^{-1}} = 0 \implies \text{diag}(\Sigma_x) = \sum_{k=1}^T x_k - f_{\Theta}(x_{k-1}, u_k)$$