Exercise Sheet 8

Machine Learning 2, SS16

June 21, 2016

Mario Tambos, 380599; Viktor Jeney, 348969; Sascha Huk, 321249; Jan Tinapp, 0380549

Exercise 1

(a) see ipynb file

(b) Since k is a kernel we find $k(x, y) = \langle\Phi(x), \Phi(y)\rangle$ for all sequences x and y.

We have to show that $v^\top Kv > 0$ for all $v = (v_1, \dots, v_n)$. More precisely, for all the sequences x_i (for all $i = 1, \dots, n$) we have to show:

$$\sum_{k,l=1}^{n} v_k v_l k(x_k, x_l) \geq 0$$

$$\iff \sum_{k,l=1}^{n} v_k v_l \Big\langle \Phi(x_k), \Phi(x_l) \Big\rangle \geq 0$$

$$\iff \sum_{k,l=1}^{n} v_l \Big\langle v_k \Phi(x_k), \Phi(x_l) \Big\rangle \geq 0$$

$$\iff \sum_{k,l=1}^{n} \Big\langle v_k \Phi(x_k), v_l \Phi(x_l) \Big\rangle \geq 0$$

$$\iff \sum_{l=1}^{n} \Big\langle \sum_{k=1}^{n} v_k \Phi(x_k), v_l \Phi(x_l) \Big\rangle \geq 0$$

$$\iff \Big\langle \sum_{k=1}^{n} v_k \Phi(x_k), \sum_{l=1}^{n} v_l \Phi(x_l) \Big\rangle \geq 0$$

$$\iff \left\| \sum_{k=1}^{n} v_k \Phi(x_k) \right\|^2 \geq 0$$

We only can conjecture $\forall x.\Phi(x) = x * x$. We try to proof it and compare $\Big\langle \Phi(x), \Phi(x') \Big\rangle$ with $\|x * x'\|$:

$$\Big\langle \Phi(x), \Phi(x') \Big\rangle = \sum_i (x * x)_i (x' * x')_i = \sum_i (\sum_{k=1}^{i} x_k x_{i-k})(\sum_{k=1}^{i} x'_k x'_{i-k}) = \sum_i \sum_{k=1}^{i} (x_k x_{i-k}) * (x'_k x'_{i-k})$$

$$\|x * x'\| = \sum_i (x * x')_i (x' * x)_i = \sum_i (\sum_{k=1}^{i} x_k x'_{i-k})(\sum_{k=1}^{i} x'_k x_{i-k}) = \sum_i \sum_{k=1}^{i} (x_k x'_{i-k}) * (x'_k x_{i-k})$$

We are not able to compute further on the rush. Conjecturally, the terms are equal.

Exercise 2

$$y_t = (w * x)_t = \sum_{s \in \mathbb{Z}} w_s x_{t-s}$$

(a)

$$\begin{aligned}
\frac{\partial E}{\partial x_k} &= \sum_{t \in \mathbb{Z}} \frac{\partial E}{\partial y_t} \frac{\partial y_t}{\partial x_k} \\
&= \sum_{t \in \mathbb{Z}} \frac{\partial E}{\partial y_t} \frac{\partial}{\partial x_k} (w * x)_t \\
&= \sum_{t \in \mathbb{Z}} \frac{\partial E}{\partial y_t} \sum_{s \in \mathbb{Z}} w_s \frac{\partial}{\partial x_k} x_{t-s} \\
&= \sum_{t \in \mathbb{Z}} \frac{\partial E}{\partial y_t} \sum_{s \in \mathbb{Z}} w_s \mathbb{1}_{\{t-s=k\}} \\
&= \sum_{t \in \mathbb{Z}} \frac{\partial E}{\partial y_t} w_{-k+t} \\
&= \left[\frac{\partial E}{\partial y} \star w \right]_{-k}
\end{aligned}$$

(b)

$$\begin{aligned}
\frac{\partial E}{\partial w_k} &= \sum_{t \in \mathbb{Z}} \frac{\partial E}{\partial y_t} \frac{\partial y_t}{\partial w_k} \\
&= \sum_{t \in \mathbb{Z}} \frac{\partial E}{\partial y_t} \frac{\partial}{\partial w_k} (w * x)_t \\
&= \sum_{t \in \mathbb{Z}} \frac{\partial E}{\partial y_t} \sum_{s \in \mathbb{Z}} x_{t-s} \frac{\partial}{\partial w_k} w_s \\
&= \sum_{t \in \mathbb{Z}} \frac{\partial E}{\partial y_t} \sum_{s \in \mathbb{Z}} x_{t-s} \mathbb{1}_{\{s=k\}} \\
&= \sum_{t \in \mathbb{Z}} \frac{\partial E}{\partial y_t} x_{-k+t} \\
&= \left[\frac{\partial E}{\partial y} \star x \right]_{-k}
\end{aligned}$$

Exercise 3

(a)

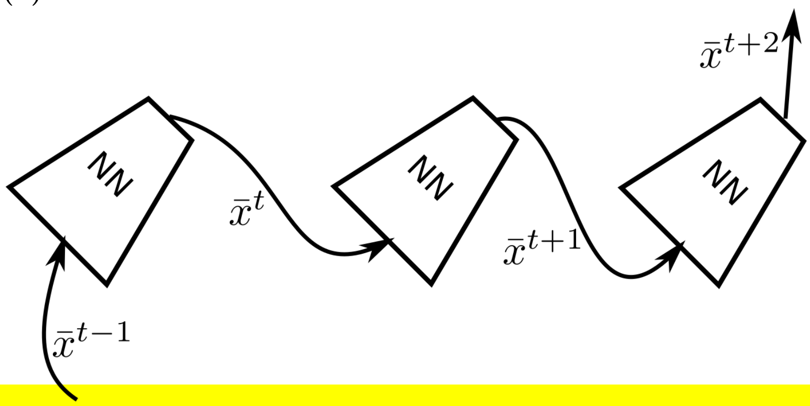
Euler discretization gives us for every $j = 1, .., d$:

$$x_j^t - x_j^{t-1} = 0.1(\tanh(\sum_{i=1}^{d} x_i^{t-1} w_{ij} + b_j) - x_j^{t-1})$$

The transition function is then component wise defined as

$$\Theta(x)_j = 0.1(\tanh(\sum_{i=1}^{d} x_i w_{ij} + b_j) - x_j) + x_j$$

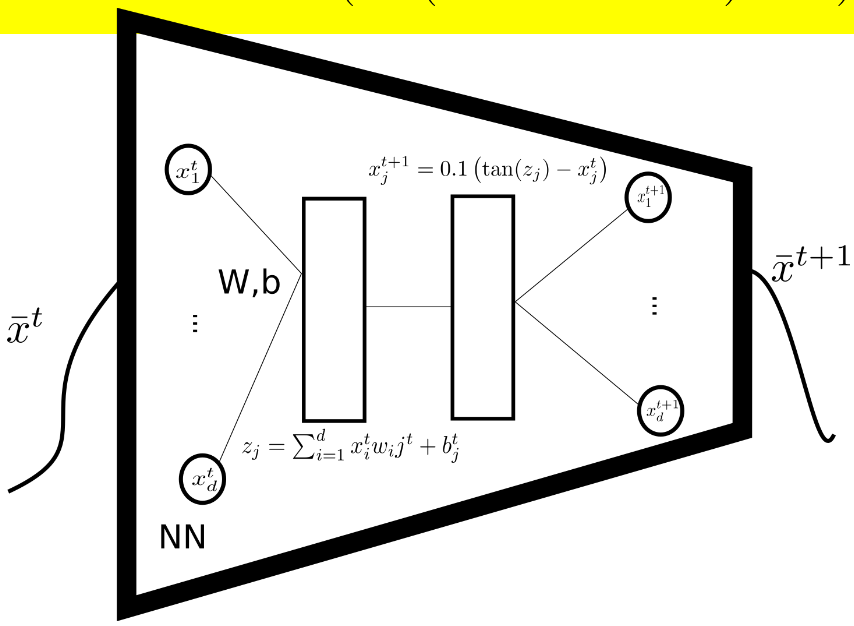
(b)



$\bar{x}^t = [x_1^t, \ldots, x_d^t]^T$

$\forall j \in (1, d) x_j^0 \rightarrow$ initial input

$\forall j \in (1, d), t > 0 x_j^t = 0.1 \left(\tanh\left(\sum_{i=1}^d x_i^{t-1} w_{ij}^{t-1} - b_j^{t-1} \right) - x_j^{t-1} \right)$



(c)
if you

$$\begin{aligned}
\frac{\partial x_i^t}{\partial x_j^{t-1}} &= 0.1 \frac{\partial}{\partial x_j^{t-1}} (\tanh(\sum_{k=1}^d x_k^{t-1} w_{ki} + b_i) - x_i^{t-1}) + \frac{\partial x_i^{t-1}}{\partial x_j^{t-1}} \\
&= 0.1 \frac{\partial}{\partial x_j^{t-1}} (\tanh(\sum_{k=1}^d x_k^{t-1} w_{ki} + b_i)) + 0.9 \mathbb{1}_{\{i=j\}} \\
&= 0.1 (1 - \tanh(\sum_{k=1}^d x_k^{t-1} w_{ki} + b_i)) \sum_{k=1}^d \frac{\partial}{\partial x_j^{t-1}} x_k^{t-1} w_{ki} + 0.9 \mathbb{1}_{\{i=j\}} \\
&= 0.1 (1 - \tanh(\sum_{k=1}^d x_k^{t-1} w_{ki} + b_i)) \sum_{k=1}^d \mathbb{1}_{\{k=j\}} w_{ki} + 0.9 \mathbb{1}_{\{i=j\}} \\
&= 0.1 (1 - \tanh(\sum_{k=1}^d x_k^{t-1} w_{ki} + b_i)) w_{ji} + 0.9 \mathbb{1}_{\{i=j\}}
\end{aligned}$$

(d)

$$\begin{aligned}
\frac{\partial x_i^t}{\partial b_j} &= 0.1 \frac{\partial}{\partial b_j} \tanh(\sum_{k=1}^{d} x_k^{t-1} w_{ki} + b_i) \\
&= 0.1(1 - \tanh(\sum_{k=1}^{d} x_k^{t-1} w_{ki} + b_i)) \frac{\partial b_i}{\partial b_j} \\
&= 0.1(1 - \tanh(\sum_{k=1}^{d} x_k^{t-1} w_{ki} + b_i)) \mathbb{1}_{\{i=j\}}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial x_i^t}{\partial w_{jl}} &= 0.1 \frac{\partial}{\partial w_{jl}} \tanh(\sum_{k=1}^{d} x_k^{t-1} w_{ki} + b_i) \\
&= 0.1(1 - \tanh(\sum_{k=1}^{d} x_k^{t-1} w_{ki} + b_i)) \sum_{k=1}^{d} x_k^{t-1} \frac{\partial}{\partial w_{jl}} w_{ki} \\
&= 0.1(1 - \tanh(\sum_{k=1}^{d} x_k^{t-1} w_{ki} + b_i)) \sum_{k=1}^{d} x_k^{t-1} \mathbb{1}_{\{j=k\}} \mathbb{1}_{\{l=i\}} \\
&= 0.1(1 - \tanh(\sum_{k=1}^{d} x_k^{t-1} w_{ki} + b_i)) x_j^{t-1} \mathbb{1}_{\{l=i\}}
\end{aligned}$$