

000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046
047
048
049
050
051
052
053

Supplementary Material for the RNN-RNADE model

Anonymous Author(s)
Affiliation
Address
email

1 Introduction

This document contains derivations of the gradients for the RNN-RNADE. The cost function for training the RNN-RNADE is given by:

054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107

Algorithm 1 RNADE gradients

```

a  $\leftarrow$  c
for  $d$  from 1 to  $D$  do
    a  $\leftarrow$  a +  $x_d \mathbf{W}_{:,d}$ 
end for
for  $d$  from  $D$  to 1 do
     $\psi \leftarrow \rho_d \mathbf{a}$ 
     $\mathbf{h} \leftarrow \sigma(\psi)$ 
     $\mathbf{z}^\alpha \leftarrow \mathbf{V}_d^{\alpha T} \mathbf{h} + \mathbf{b}_d^\alpha$ 
     $\mathbf{z}^\mu \leftarrow \mathbf{V}_d^{\mu T} \mathbf{h} + \mathbf{b}_d^\mu$ 
     $\mathbf{z}^\sigma \leftarrow \mathbf{V}_d^{\sigma T} \mathbf{h} + \mathbf{b}_d^\sigma$ 
     $\alpha \leftarrow \text{softmax}(\mathbf{z}^\alpha)$ 
     $\mu = \mathbf{z}^\mu$ 
     $\sigma \leftarrow \exp(\mathbf{z}^\sigma)$ 
     $\phi \leftarrow \frac{1}{2} \frac{(\mu - \mathbf{x}_d)^2}{\sigma^2} - \log \sigma - \frac{1}{2} \log(2\pi)$ 
     $\pi \leftarrow \frac{\alpha \phi}{\sum_{j=1}^K \alpha_j \phi_j}$ 
     $\partial z^\alpha \leftarrow \pi - \alpha$ 
     $\partial \mathbf{V}_d^\alpha \leftarrow \partial z^\alpha \mathbf{h}$ 
     $\partial \mathbf{b}_d^\alpha \leftarrow \partial z^\alpha$ 
     $\partial z^\mu \leftarrow \pi (x_d - \mu) / \sigma^2$ 
     $\partial \mathbf{V}_d^\mu \leftarrow \partial z^\mu \mathbf{h}$ 
     $\partial \mathbf{b}_d^\mu \leftarrow \partial z^\mu$ 
     $\partial z^\sigma \leftarrow \pi \{ (x_d - \mu) / \sigma^2 - 1 \}$ 
     $\partial \mathbf{V}_d^\sigma \leftarrow \partial z^\sigma \mathbf{h}$ 
     $\partial \mathbf{b}_d^\sigma \leftarrow \partial z^\sigma$ 
     $\partial \mathbf{h} \leftarrow z^\alpha \mathbf{V}_d^\alpha + z^\mu \mathbf{V}_d^\mu + z^\sigma \mathbf{V}_d^\sigma$ 
     $\partial \phi = \partial \mathbf{h} \sigma(\psi) (1 - \sigma(\psi))$ 
     $\partial \rho_d \leftarrow \sum_j \partial \psi_j a_j$ 
     $\partial \mathbf{a} \leftarrow \partial \mathbf{a} + \partial \psi_\rho$ 
     $\partial \mathbf{W}_{:,d} \leftarrow \partial \mathbf{a} x_d$ 
    if  $d = 1$  then
         $\partial \mathbf{c} \leftarrow \partial \mathbf{a}$ 
    else
         $\mathbf{a} \leftarrow \mathbf{a} - x_d \mathbf{W}_{:,d}$ 
    end if
end for

```

Algorithm 2 RNN-RNADE gradients

```

for  $t$  from  $T$  to 1 do
   $\mathbf{a} \leftarrow \mathbf{c}$ 
  for  $d$  from 1 to  $D$  do
     $\mathbf{a} \leftarrow \mathbf{a} + x_d \mathbf{W}_{:,d}$ 
  end for
  for  $d$  from  $D$  to 1 do
     $\psi_t \leftarrow \rho_d \mathbf{a}$ 
     $\mathbf{h}_t \leftarrow \sigma(\psi_t)$ 
     $\mathbf{z}_t^\alpha \leftarrow \mathbf{V}_d^{\alpha T} \mathbf{h}_t + \mathbf{b}_{d(t)}^\alpha$ 
     $\mathbf{z}_t^\mu \leftarrow \mathbf{V}_d^{\mu T} \mathbf{h}_t + \mathbf{b}_{d(t)}^\mu$ 
     $\mathbf{z}_t^\sigma \leftarrow \mathbf{V}_d^{\sigma T} \mathbf{h}_t + \mathbf{b}_{d(t)}^\sigma$ 
     $\boldsymbol{\alpha}_t \leftarrow \text{softmax}(\mathbf{z}_t^\alpha)$ 
     $\boldsymbol{\mu}_t = \mathbf{z}_t^\mu$ 
     $\boldsymbol{\sigma}_t \leftarrow \exp(\mathbf{z}_t^\sigma)$ 
     $\phi_t \leftarrow \frac{1}{2} \frac{(\boldsymbol{\mu}_t - \mathbf{x}_d^t)^2}{\boldsymbol{\sigma}_t^2} - \log \boldsymbol{\sigma}_t - \frac{1}{2} \log(2\pi)$ 
     $\boldsymbol{\pi}_t \leftarrow \frac{\boldsymbol{\alpha}_t \phi_t}{\sum_{j=1}^K \alpha_j \phi_j}$ 
     $\partial z_t^\alpha \leftarrow \boldsymbol{\pi}_t - \boldsymbol{\alpha}_t$ 
     $\partial \mathbf{V}_d^\alpha \leftarrow \partial z_t^\alpha \mathbf{h}_t$ 
     $\partial \mathbf{b}_{d(t)}^\alpha \leftarrow \partial z_t^\alpha$ 
     $\partial z_t^\mu \leftarrow \boldsymbol{\pi}_t (x_d - \boldsymbol{\mu}_t) / \boldsymbol{\sigma}_t^2$ 
     $\partial \mathbf{V}_d^\mu \leftarrow \partial z_t^\mu \mathbf{h}_t$ 
     $\partial \mathbf{b}_{d(t)}^\mu \leftarrow \partial z_t^\mu$ 
     $\partial z_t^\sigma \leftarrow \boldsymbol{\pi}_t \{ (x_d - \boldsymbol{\mu}_t) / \boldsymbol{\sigma}_t^2 - 1 \}$ 
     $\partial \mathbf{V}_d^\sigma \leftarrow \partial z_t^\sigma \mathbf{h}_t$ 
     $\partial \mathbf{b}_{d(t)}^\sigma \leftarrow \partial z_t^\sigma$ 
     $\partial \mathbf{h} \leftarrow z_t^\alpha \mathbf{V}_d^\alpha + z_t^\mu \mathbf{V}_d^\mu + z_t^\sigma \mathbf{V}_d^\sigma$ 
     $\partial \phi_t = \partial \mathbf{h}_t \sigma(\psi_t) (1 - \sigma(\psi_t))$ 
     $\partial \rho_d(t) \leftarrow \sum_j \partial \psi_j a_j$ 
     $\partial \mathbf{a} \leftarrow \partial \mathbf{a} + \partial \psi_\rho$ 
     $\partial \mathbf{W}_{:,d} \leftarrow \partial \mathbf{a} x_d$ 
    if  $d = 1$  then
       $\partial \mathbf{c} \leftarrow \partial \mathbf{a}$ 
    else
       $\mathbf{a} \leftarrow \mathbf{a} - x_d^t \mathbf{W}_{:,d}$ 
    end if
  end for
   $\partial W_\alpha \leftarrow \partial W_\alpha + \partial \mathbf{b}_t^\alpha \mathbf{h}^{t-1 T}$ 
   $\partial W_\mu \leftarrow \partial W_\mu + \partial \mathbf{b}_t^\mu \mathbf{h}^{t-1 T}$ 
   $\partial W_\sigma \leftarrow \partial W_\sigma + \partial \mathbf{b}_t^\sigma \mathbf{h}^{t-1 T}$ 
   $\partial \mathbf{h}^t \leftarrow W_{rec} \partial \mathbf{h}^{t+1} \mathbf{h}^{t+1} (1 - \mathbf{h}^{t+1}) + W_\alpha \partial \mathbf{b}_{t+1}^\alpha + W_\mu \partial \mathbf{b}_{t+1}^\mu + W_\sigma \partial \mathbf{b}_{t+1}^\sigma$ 
   $\partial \mathbf{b}_h \leftarrow \partial \mathbf{b}_h + \partial \mathbf{h}^t \mathbf{h}^t (1 - \mathbf{h}^t)$ 
   $\partial W_{rec} \leftarrow \partial W_{rec} + \partial \mathbf{h}^t \mathbf{h}^t (1 - \mathbf{h}^t) \mathbf{h}^{t-1 T}$ 
   $\partial W_{in} \leftarrow \partial W_{in} + \partial \mathbf{h}^t \mathbf{h}^t (1 - \mathbf{h}^t) \mathbf{x}_t^T$ 
end for

```
