Practical-1: Pen and paper exercises

Georgios Methenitis

November 8, 2016

Exercise 1

We start by the loss function which is given by $\mathcal{L} = 0.5(y_{out} - y_{gt})^2$. Using this we compute the derivatives of the loss function with respect to the weights starting from

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial W_{out}} =& \frac{\partial}{\partial W_{out}} \left(0.5(y_{out} - y_{gt})^2\right) \\
=& 2 \times 0.5(y_{out} - y_{gt}) \frac{\partial}{\partial W_{out}} (y_{out} - y_{gt}) \\
=& (y_{out} - y_{gt}) \frac{\partial}{\partial W_{out}} y_{out} \\
=& (y_{out} - y_{gt}) \frac{\partial}{\partial W_{out}} \left(f_3(W_{out} f_2(w_2 f_1(w_1 x_{in})))\right), \text{by the chain rule} \\
=& (y_{out} - y_{gt}) \frac{\partial f_3(s_{out})}{\partial W_{out}} z_2.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial W_2} =& (y_{out} - y_{gt}) \frac{\partial \mathcal{L}}{\partial W_2} \left(f_3(W_{out} f_2(w_2 f_1(w_1 x_{in})))\right), \text{by the chain rule} \\
=& (y_{out} - y_{gt}) \frac{\partial f_3(s_{out})}{\partial W_{out}} \frac{\partial \mathcal{L}}{\partial W_2} \left(f_2(w_2 f_1(w_1 x_{in}))\right) \\
&= (y_{out} - y_{gt}) \frac{\partial f_3(s_{out})}{\partial W_{out}} \frac{\partial f_2(s_2)}{\partial W_2} z_1.
\end{aligned}$$

And similarly,

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial W_1} =& (y_{out} - y_{gt}) \frac{\partial \mathcal{L}}{\partial W_1} \left(f_3(W_{out} f_2(w_2 f_1(w_1 x_{in})))\right), \text{by the chain rule} \\
&= (y_{out} - y_{gt}) \frac{\partial f_3(s_{out})}{\partial W_{out}} \frac{\partial f_2(s_2)}{\partial W_2} \frac{\partial \mathcal{L}}{\partial W_1} \left(f_1(w_1 x_{in})\right)) \\
&= (y_{out} - y_{gt}) \frac{\partial f_3(s_{out})}{\partial W_{out}} \frac{\partial f_2(s_2)}{\partial W_2} \frac{\partial f_1(s)}{\partial W_1}.
\end{aligned}$$

Prelude

We start by $\Delta W_N = \dfrac{\partial \mathcal{L}}{\partial W_N}$,

$$\begin{aligned}
\Delta W_N =& \frac{\partial \mathcal{L}}{\partial W_N} = \delta_N \\
\Delta W_{N-1} =& \delta_N w_{N-1} \frac{\partial \mathcal{L}}{\partial W_{N-1}} \\
\Delta W_{N-2} =& \delta_{N-1} w_{N-2} \frac{\partial \mathcal{L}}{\partial W_{N-2}} \\
&\vdots \\
\Delta W_0 =& \delta_1 w_0 \frac{\partial \mathcal{L}}{\partial W_0}
\end{aligned}$$

We can write the general form for the weight updates $\Delta W_{i \to j} = \delta_j z_i$.

Exercise-2

Iteration #1 for the first sample $[0.75, 0.8]$:

- $s_i = [0.458, 0.869, 0.704]$
- $z_i = [0.458, 0.869, 0.704]$
- $s_{out} = 0.09858999999999997$
- $z_{out} = 0.09858999999999997$, using ReLU as activation function in the output unit.
- $\mathcal{L} = 0.40626999405000003$
- $\delta_{out} = (z_{out} - x[0]) \dfrac{\partial \mathcal{L}}{\partial s_{out}} = -0.09858999999999997$
- $\delta_i = \delta_{out} \dfrac{\partial \mathcal{L}}{\partial s_i} = [-0.90141, -0.90141, -0.90141]$
- $w = [0.10256916, 0.18666506, 0.21691853]$
- $W = [[0.7352115, 0.8352115, 0.1352115], [0.1542256, 0.5742256, 1.0242256]]$

Using the following script in python we proceed:

```
_x = x[0] # sample
_y = y[0] # target
s_i = np.dot(_x, W) # forward propagation
z_i = relu(s_i) # relu at s_i
s_out = np.dot(z_i, w) # forward propagation
```

```
z_out = relu(s_out) # relu at s_out
L = error(z_out, _y) # error
delta_out = (z_out - _y) * d_relu(s_out) # error signal at s_out
delta_i = delta_out * d_relu(s_i) # error signal at s_i
Delta_w = - theta * delta_out * z_i # derivative at weights w
Delta_W = - theta * delta_i * _x.reshape((2,1)) # derivative at weights W
w = w + Delta_w # new w
W = W + Delta_W # new W
```

Here we present the results for all iterations and all samples:

```
-Iteration:  0
---- sample:  0
     _x:  [ 0.75  0.8 ]
     _y:  1
     s_i:  [ 0.458  0.869  0.704]
     z_i:  [ 0.458  0.869  0.704]
     s_out:  0.09859
     z_out:  0.09859
     L:  0.40626999405
     delta_out:  -0.90141
     delta_i:  [-0.90141 -0.90141 -0.90141]
     Delta_w:  [ 0.08256916  0.15666506  0.12691853]
     Delta_W:  [[ 0.1352115  0.1352115  0.1352115]
 [ 0.1442256  0.1442256  0.1442256]]
     w:  [ 0.10256916  0.18666506  0.21691853]
     W:  [[ 0.7352115  0.8352115  0.1352115]
 [ 0.1542256  0.5742256  1.0242256]]

---- sample:  1
     _x:  [ 0.2   0.05]
     _y:  1
     s_i:  [ 0.15475358  0.19575358  0.07825358]
     z_i:  [ 0.15475358  0.19575358  0.07825358]
     s_out:  0.0693879488373
     z_out:  0.0693879488373
     L:  0.433019394885
     delta_out:  -0.930612051163
     delta_i:  [-0.93061205 -0.93061205 -0.93061205]
     Delta_w:  [ 0.02880311  0.03643413  0.01456474]
     Delta_W:  [[ 0.03722448  0.03722448  0.03722448]
 [ 0.00930612  0.00930612  0.00930612]]
     w:  [ 0.13137227  0.22309919  0.23148327]
     W:  [[ 0.77243598  0.87243598  0.17243598]
 [ 0.16353172  0.58353172  1.03353172]]

---- sample:  2
     _x:  [-0.75  0.8 ]
     _y:  -1
     s_i:  [-0.44850161 -0.18750161  0.69749839]
     z_i:  [-0.         -0.          0.69749839]
     s_out:  0.161459210144
     z_out:  0.161459210144
     L:  0.674493748414
     delta_out:  1.16145921014
     delta_i:  [ 0.          0.          1.16145921]
     Delta_w:  [ 0.          0.         -0.16202319]
     Delta_W:  [[ 0.          0.          0.17421888]
 [-0.         -0.         -0.18583347]]
     w:  [ 0.13137227  0.22309919  0.06946009]
     W:  [[ 0.77243598  0.87243598  0.34665486]
 [ 0.16353172  0.58353172  0.84769825]]

---- sample:  3
     _x:  [ 0.2  -0.05]
     _y:  -1
     s_i:  [ 0.14631061  0.14531061  0.02694606]
     z_i:  [ 0.14631061  0.14531061  0.02694606]
     s_out:  0.053511510936
     z_out:  0.053511510936
     L:  0.554943251837
     delta_out:  1.05351151094
     delta_i:  [ 1.05351151  1.05351151  1.05351151]
     Delta_w:  [-0.03082798 -0.03061728 -0.0056776 ]
     Delta_W:  [[-0.04214046 -0.04214046 -0.04214046]
 [ 0.01053512  0.01053512  0.01053512]]
     w:  [ 0.10054428  0.19248191  0.06378249]
     W:  [[ 0.73029552  0.83029552  0.3045144 ]
 [ 0.17406684  0.59406684  0.85823336]]
```

```
-Iteration:  1
---- sample:  0
     _x:  [ 0.75  0.8 ]
     _y:  1
     s_i:  [ 0.68697511  1.09797511  0.91497249]
     z_i:  [ 0.68697511  1.09797511  0.91497249]
     s_out:  0.338770985555
     z_out:  0.338770985555
     L:  0.218611904772
     delta_out:  -0.661229014445
     delta_i:  [-0.66122901 -0.66122901 -0.66122901]
     Delta_w:  [ 0.09084957  0.1452026   0.12100127]
     Delta_W:  [[ 0.09918435  0.09918435  0.09918435]
 [ 0.10579664  0.10579664  0.10579664]]
     w:  [ 0.19139386  0.33768451  0.18478376]
     W:  [[ 0.82947987  0.92947987  0.40369876]
 [ 0.27986348  0.69986348  0.96403   ]]

---- sample:  1
     _x:  [ 0.2   0.05]
     _y:  1
     s_i:  [ 0.17988915  0.22088915  0.12894125]
     z_i:  [ 0.17988915  0.22088915  0.12894125]
     s_out:  0.132846770649
     z_out:  0.132846770649
     L:  0.375977361587
     delta_out:  -0.867153229351
     delta_i:  [-0.86715323 -0.86715323 -0.86715323]
     Delta_w:  [ 0.03119829  0.03830895  0.02236236]
     Delta_W:  [[ 0.03468613  0.03468613  0.03468613]
 [ 0.00867153  0.00867153  0.00867153]]
     w:  [ 0.22259215  0.37599345  0.20714613]
     W:  [[ 0.864166    0.964166    0.43838488]
 [ 0.28853501  0.70853501  0.97270154]]

---- sample:  2
     _x:  [-0.75  0.8 ]
     _y:  -1
     s_i:  [-0.41729649 -0.15629649  0.44937257]
     z_i:  [-0.         -0.          0.44937257]
     s_out:  0.0930857863762
     z_out:  0.0930857863762
     L:  0.597418268189
     delta_out:  1.09308578638
     delta_i:  [ 0.          0.          1.09308579]
     Delta_w:  [ 0.          0.         -0.09824055]
     Delta_W:  [[ 0.          0.          0.16396287]
 [-0.         -0.         -0.17489373]]
     w:  [ 0.22259215  0.37599345  0.10890557]
     W:  [[ 0.864166    0.964166    0.60234775]
 [ 0.28853501  0.70853501  0.79780781]]

---- sample:  3
     _x:  [ 0.2  -0.05]
     _y:  -1
     s_i:  [ 0.15840645  0.15740645  0.08057916]
     z_i:  [ 0.15840645  0.15740645  0.08057916]
     s_out:  0.103219346565
     z_out:  0.103219346565
     L:  0.608546463317
     delta_out:  1.10321934656
     delta_i:  [ 1.10321935  1.10321935  1.10321935]
     Delta_w:  [-0.03495141 -0.03473077 -0.0177793 ]
     Delta_W:  [[-0.04412877 -0.04412877 -0.04412877]
 [ 0.01103219  0.01103219  0.01103219]]
     w:  [ 0.18764074  0.34126269  0.09112628]
     W:  [[ 0.82003723  0.92003723  0.55821898]
 [ 0.2995672   0.7195672   0.80884   ]]
```

```
-Iteration:  2
---- sample:  0
     _x:  [ 0.75  0.8 ]
     _y:  1
     s_i:  [ 0.85468168  1.26568168  1.06573624]
     z_i:  [ 0.85468168  1.26568168  1.06573624]
     s_out:  0.689419606272
     z_out:  0.689419606272
     L:  0.0482300904842
     delta_out:  -0.310580393728
     delta_i:  [-0.31058039 -0.31058039 -0.31058039]
     Delta_w:  [ 0.05308947  0.07861918  0.06619936]
     Delta_W:  [[ 0.04658706  0.04658706  0.04658706]
 [ 0.04969286  0.04969286  0.04969286]]
     w:  [ 0.24073021  0.41988187  0.15732563]
     W:  [[ 0.86662429  0.96662429  0.60480604]
 [ 0.34926007  0.76926007  0.85853287]]

---- sample:  1
     _x:  [ 0.2   0.05]
     _y:  1
     s_i:  [ 0.19078786  0.23178786  0.16388785]
     z_i:  [ 0.19078786  0.23178786  0.16388785]
     s_out:  0.169035682078
     z_out:  0.169035682078
     L:  0.34525084883
     delta_out:  -0.830964317922
     delta_i:  [-0.83096432 -0.83096432 -0.83096432]
     Delta_w:  [ 0.03170758  0.03852149  0.02723699]
     Delta_W:  [[ 0.03323857  0.03323857  0.03323857]
 [ 0.00830964  0.00830964  0.00830964]]
     w:  [ 0.27243779  0.45840336  0.18456262]
     W:  [[ 0.89986286  0.99986286  0.63804461]
 [ 0.35756971  0.77756971  0.86684251]]

---- sample:  2
     _x:  [-0.75  0.8 ]
     _y:  -1
     s_i:  [-0.38884138 -0.12784138  0.21494055]
     z_i:  [-0.         -0.          0.21494055]
     s_out:  0.0396699918483
     z_out:  0.0396699918483
     L:  0.540456845975
     delta_out:  1.03966999185
     delta_i:  [ 0.          0.          1.03966999]
     Delta_w:  [ 0.          0.         -0.04469345]
     Delta_W:  [[ 0.          0.          0.1559505]
 [-0.         -0.         -0.1663472]]
     w:  [ 0.27243779  0.45840336  0.13986918]
     W:  [[ 0.89986286  0.99986286  0.79399511]
 [ 0.35756971  0.77756971  0.70049531]]

---- sample:  3
     _x:  [ 0.2  -0.05]
     _y:  -1
     s_i:  [ 0.16209409  0.16109409  0.12377426]
     z_i:  [ 0.16209409  0.16109409  0.12377426]
     s_out:  0.135318828438
     z_out:  0.135318828438
     L:  0.644474421103
     delta_out:  1.13531882844
     delta_i:  [ 1.13531883  1.13531883  1.13531883]
     Delta_w:  [-0.03680569 -0.03657863 -0.02810465]
     Delta_W:  [[-0.04541275 -0.04541275 -0.04541275]
 [ 0.01135319  0.01135319  0.01135319]]
     w:  [ 0.2356321   0.42182473  0.11176453]
     W:  [[ 0.85445011  0.95445011  0.74858236]
 [ 0.3689229   0.7889229   0.7118485 ]]
```

Exercise-3

i)

Since $\max(0, p_j - p_{y_i} + margin)$ is minimized when $p_j - p_{y_i} = -margin$ which results $p_j = p_{y_i} - margin$, the loss function is trying to maximize the difference of the probability output for the class p_{y_i} with regards to every other class j by the value of $margin$. In simple words, $\mathcal{L}_{hinge}$ is trying to maximize the difference between the correct all other classes.

ii)

$$\frac{\partial \mathcal{L}_{hinge}}{\partial o_j} = \frac{\partial}{\partial o_j}(\max(0, p_j - p_{y_i} + margin))$$

For $p_{y_i} > p_j + margin$, $\frac{\partial \mathcal{L}_{hinge}}{\partial o_j} = 0$, while for $p_{y_i} = p_j + margin$, $\frac{\partial \mathcal{L}_{hinge}}{\partial o_j} = \varnothing$. We assume $p_{y_i} < p_j + margin$,

$$\begin{aligned}
\frac{\partial \mathcal{L}_{hinge}}{\partial o_j} =& \frac{\partial}{\partial o_j}(p_j - p_{y_i} + margin)) \\
=& \frac{\partial}{\partial o_j} p_j \\
=& \frac{\partial}{\partial o_j}\left(\frac{\exp(o_j)}{\sum_k \exp(o_k)}\right) \\
=& \frac{\partial}{\partial o_j}\left(\exp(o_j)\frac{1}{\sum_k \exp(o_k)}\right) \\
=& (\exp(o_j))' \frac{1}{\sum_k \exp(o_k)} + \exp(o_j)\left(\frac{1}{\sum_k \exp(o_k)}\right)' \\
=& \exp(o_j)\frac{1}{\sum_k \exp(o_k)} - \exp(o_j)\left(\frac{1}{\sum_k \exp(o_k)}\right)^2 \left(\sum_k \exp(o_j)\right)' \\
=& \exp(o_j)\frac{1}{\sum_k \exp(o_k)} - \exp(2o_j)\left(\frac{1}{\sum_k \exp(o_k)}\right)^2 \\
=& p_j - p_j^2
\end{aligned}$$

which is the derivative of the loss function $\mathcal{L}_{hinge}$ with respect to o_j.

Exercise-4