Practical 1

Vasileios Charatsidis

July 1, 2018

1 Exersise 1

1.1 Compute (generalized) derivatives of the loss function, with respect to the weights.

$$\mathcal{L} = 0.5 * (y_{out} - y_{gt})^2 \tag{1}$$

$$\textbf{for } \boldsymbol{\frac{\partial \mathcal{L}}{\partial W_{out}}}$$

$$\frac{\partial \mathcal{L}}{\partial W_{out}} = \frac{\partial \mathcal{L}}{\partial y_{out}} \cdot \frac{\partial y_{out}}{\partial s_{out}} \cdot \frac{\partial s_{out}}{\partial W_{out}} \Rightarrow \tag{2}$$

$$\frac{\partial \mathcal{L}}{\partial W_{out}} = (y_{out} - s_{out}) \cdot \frac{\partial f_3(s_{out})}{\partial s_{out}} \cdot \frac{\partial (W_{out} \cdot z_2)}{\partial W_{out}} \Rightarrow$$

$$\frac{\partial \mathcal{L}}{\partial W_{out}} = (y_{out} - s_{out}) \cdot f\,'_3(s_{out}) \cdot z_2$$

$$\textbf{for } \boldsymbol{\frac{\partial \mathcal{L}}{\partial W_2}}$$

$$\frac{\partial \mathcal{L}}{\partial W_2} = \frac{\partial \mathcal{L}}{\partial y_{out}} \cdot \frac{\partial y_{out}}{\partial s_{out}} \cdot \frac{\partial s_{out}}{\partial z_2} \cdot \frac{\partial z_2}{\partial s_2} \cdot \frac{\partial s_2}{\partial W_2} \Rightarrow \tag{3}$$

$$\frac{\partial \mathcal{L}}{\partial W_2} = (y_{out} - s_{out}) \cdot f\,'_3(s_{out}) \cdot W_{out} \cdot f\,'_2(s_2) \cdot z_1$$

$$\textbf{similarly for } \boldsymbol{\frac{\partial \mathcal{L}}{\partial W_1}}$$

$$\frac{\partial \mathcal{L}}{\partial W_1} = \frac{\partial \mathcal{L}}{\partial y_{out}} \cdot \frac{\partial y_{out}}{\partial s_{out}} \cdot \frac{\partial s_{out}}{\partial z_2} \cdot \frac{\partial z_2}{\partial s_2} \cdot \frac{\partial s_2}{\partial z_1} \cdot \frac{\partial z_1}{\partial s_1} \cdot \frac{\partial s_1}{\partial W_1} \Rightarrow \tag{4}$$

$$\frac{\partial \mathcal{L}}{\partial W_1} = (y_{out} - s_{out}) \cdot f\,'_3(s_{out}) \cdot W_{out} \cdot f\,'_2(s_2) \cdot W_2 \cdot f\,'_1(s_1) \cdot x_{in}$$

1.2 Write down ΔW_k in terms of δ_k

$$\text{We know that: } \delta_k = \frac{\partial \mathcal{L}}{\partial s_k} \text{ and that: } \frac{\partial \mathcal{L}}{\partial s_k} = \frac{\partial \mathcal{L}}{\partial z_k} \cdot \frac{\partial z_k}{\partial s_k}$$

$$\text{So: } \delta_k = \frac{\partial \mathcal{L}}{\partial z_k} \cdot \frac{\partial z_k}{\partial s_k} \tag{5}$$

We can rewrite equations (2), (3) and (4) as:

$$\frac{\partial \mathcal{L}}{\partial W_{out}} = \underbrace{\frac{\partial \mathcal{L}}{\partial z_{out}} \cdot \frac{\partial z_{out}}{\partial s_{out}}}_{\delta_{out}} \cdot \frac{\partial s_{out}}{\partial W_{out}} \Rightarrow$$

$$\Delta W_{out} = \frac{\partial \mathcal{L}}{\partial W_{out}} = \delta_{out} \cdot \frac{\partial s_{out}}{\partial W_{out}} = \delta_{out} \cdot z_2 \tag{6}$$

$$\frac{\partial \mathcal{L}}{\partial W_2} = \frac{\partial \mathcal{L}}{\partial z_{out}} \cdot \frac{\partial z_{out}}{\partial s_{out}} \cdot \frac{\partial s_{out}}{\partial z_2} \cdot \frac{\partial z_2}{\partial s_2} \cdot \frac{\partial s_2}{\partial W_2} \tag{7}$$

$$\delta_2 = \frac{\partial \mathcal{L}}{\partial s_2} = \frac{\partial \mathcal{L}}{\partial z_{out}} \cdot \frac{\partial z_{out}}{\partial s_{out}} \cdot \frac{\partial s_{out}}{\partial z_2} \cdot \frac{\partial z_2}{\partial s_2} \tag{8}$$

$$\text{from (7) and (8) is implied that } \Delta W_2 = \frac{\partial \mathcal{L}}{\partial W_2} = \delta_2 \cdot \frac{\partial s_2}{\partial W_2} = \delta_2 \cdot z_1 \tag{9}$$

similarly for δ_1

$$\frac{\partial \mathcal{L}}{\partial W_1} = \underbrace{\frac{\partial \mathcal{L}}{\partial z_{out}} \cdot \frac{\partial z_{out}}{\partial s_{out}} \cdot \frac{\partial s_{out}}{\partial z_2} \cdot \frac{\partial z_2}{\partial s_2} \cdot \frac{\partial s_2}{\partial z_1} \cdot \frac{\partial z_1}{\partial s_1}}_{\delta_1} \cdot \frac{\partial s_1}{\partial W_1} \Rightarrow$$

$$\text{is implied that } \Delta W_1 = \frac{\partial \mathcal{L}}{\partial W_1} = \delta_1 \cdot \frac{\partial s_1}{\partial W_1} = \delta_1 \cdot x_{in} \tag{10}$$

2 Exercise 2

$$x = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_2^{(1)} & x_4^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & x_4^{(2)} \end{bmatrix} = \begin{bmatrix} 0.75 & 0.2 & -0.75 & 0.2 \\ 0.8 & 0.05 & 0.8 & -0.05 \end{bmatrix}$$

$$W = \begin{bmatrix} W_{11} & W_{21} & W_{31} & \\ W_{12} & W_{22} & W_{32} & \end{bmatrix} = \begin{bmatrix} 0.6 & 0.7 & 0.0 \\ 0.01 & 0.43 & 0.88 \end{bmatrix}$$

$$w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 0.02 \\ 0.03 \\ 0.9 \end{bmatrix}$$

$$y = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}$$

$$\alpha = 0.05$$

$$s_1 = W^T \cdot x \tag{1}$$

$$z_1 = ReLU(s_1) \tag{2}$$

$$s_{out} = w^T \cdot z_1 \tag{3}$$

$$y_{out} = z_{out} = tanh(s_{out}) \tag{4}$$

$$\mathcal{L} = 0.5 \cdot (y_{out} - y)^2 \tag{5}$$

$$\delta_{out} = \frac{\partial \mathcal{L}}{\partial s_{out}} \tag{6}$$

$$\delta_1 = \delta_{out} \cdot w \cdot \frac{\partial f_{(s_1)}}{\partial s_1} \tag{7}$$

$$\Delta w = \delta_{out} \cdot z_1 \tag{8}$$

$$\Delta W = \delta_1 \cdot x_1 \tag{9}$$

$$\Delta W = \delta_1 \cdot x_1 \tag{10}$$

$$W = W - \alpha \cdot \Delta W \tag{11}$$

$$w = w - \alpha \cdot \Delta w \tag{12}$$

If we follow the steps from 1 to 12 with our given input we have done 1 iteration of forward and backward pass.

2.1 iteration 1

$$s_1 = \begin{bmatrix} 0.458 & 0.1205 & -0.442 & 0.1195 \\ 0.869 & 0.1615 & -0.181 & 0.1185 \\ 0.704 & 0.044 & 0.704 & -0.044 \end{bmatrix}$$

$$z_1 = \begin{bmatrix} 0.458 & 0.1205 & 0 & 0.1195 \\ 0.869 & 0.1615 & 0 & 0.1185 \\ 0.704 & 0.044 & 0.704 & 0 \end{bmatrix}$$

$$s_{out} = \begin{bmatrix} 0.09859 & 0.011215 & 0.06336 & 0.005945 \end{bmatrix}$$

$$y_{out} = z_{out} = \begin{bmatrix} 0.09827181 & 0.01121453 & 0.06327535 & 0.00594493 \end{bmatrix}$$

$$\mathcal{L} = \begin{bmatrix} 0.40655687 & 0.48884835 & 0.56527723 & 0.5059626 \end{bmatrix}$$

$$\delta_{out} = \begin{bmatrix} -0.89301989 & -0.98866111 & 1.05901824 & 1.00590938 \end{bmatrix}$$

$$\delta_1 = \begin{bmatrix} -0.0178604 & -0.01977322 & 0 & 0.02011819 \\ -0.0267906 & -0.02965983 & 0 & 0.03017728 \\ -0.08037179 & -0.0889795 & 0.09531164 & 0 \end{bmatrix}$$

$$\Delta w = \begin{bmatrix} -0.4079306 & -0.81650279 & 0.07336175 \end{bmatrix}$$

$$\Delta W = \begin{bmatrix} -0.01332631 & -0.01998946 & -0.14955847 \\ -0.01628289 & -0.02442433 & 0.00750291 \end{bmatrix}$$

$$w\,' = \begin{bmatrix} -0.4079306 & -0.81650279 & 0.07336175 \end{bmatrix}$$

$$W\,' = \begin{bmatrix} 0.60133263 & 0.70199895 & 0.01495585 \\ 0.01162829 & 0.43244243 & 0.87924971 \end{bmatrix}$$

2.2 iteration 2

$$s_1 = \begin{bmatrix} 0.460302 & 0.120848 & -0.441697 & 0.119685 \\ 0.872453 & 0.162022 & -0.180545 & 0.118778 \\ 0.714617 & 0.0469537 & 0.692183 & -0.0409713 \end{bmatrix}$$

$$z_1 = \begin{bmatrix} 0.46030 & 0.120848 & 0 & 0.119685 \\ 0.872453 & 0.162022 & 0 & 0.118778 \\ 0.714617 & 0.0469537 & 0.692183 & 0 \end{bmatrix}$$

$$s_{out} = \begin{bmatrix} 0.184466 & 0.0293179 & 0.0572185 & 0.0205376 \end{bmatrix}$$

$$y_{out} = z_{out} = \begin{bmatrix} 0.182402 & 0.0293095 & 0.0571561 & 0.0205347 \end{bmatrix}$$

$$\mathcal{L} = \begin{bmatrix} 0.334234 & 0.47112 & 0.55879 & 0.520746 \end{bmatrix}$$

$$\delta_{out} = \begin{bmatrix} -0.790397 & -0.969857 & 1.0537 & 1.0201 \end{bmatrix}$$

$$\delta_1 = \begin{bmatrix} -0.0480506 & -0.0589606 & 0 & 0.0620153 \\ -0.088248 & -0.108285 & 0 & 0.113895 \\ -0.0653372 & -0.0801721 & 0.0871031 & 0 \end{bmatrix}$$

$$\Delta w = \begin{bmatrix} -0.358935 & -0.725557 & 0.118986 \end{bmatrix}$$

$$\Delta W = \begin{bmatrix} -0.035427 & -0.065064 & -0.130365 \\ -0.0444893 & -0.0817074 & 0.0134041 \end{bmatrix}$$

$$w\,' = \begin{bmatrix} 0.0966866 & 0.184206 & 0.0707652 \end{bmatrix}$$

$$W\,' = \begin{bmatrix} 0.604875 & 0.708505 & 0.0279923 \\ 0.0160772 & 0.440613 & 0.877909 \end{bmatrix}$$

2.3 iteration 3

$$s_1 = \begin{bmatrix} 0.466518 & 0.121779 & -0.440795 & 0.120171 \\ 0.88387 & 0.163732 & -0.178888 & 0.11967 \\ 0.723322 & 0.0494939 & 0.681333 & -0.038297 \end{bmatrix}$$

$$z_1 = \begin{bmatrix} 0.466518 & 0.121779 & 0 & 0.120171 \\ 0.88387 & 0.163732 & 0 & 0.11967 \\ 0.723322 & 0.0494939 & 0.681333 & 0 \end{bmatrix}$$

$$s_{out} = \begin{bmatrix} 0.259106 & 0.0454372 & 0.0482147 & 0.0336629 \end{bmatrix}$$

$$y_{out} = z_{out} = \begin{bmatrix} 0.253459 & 0.0454059 & 0.0481774 & 0.0336502 \end{bmatrix}$$

$$\mathcal{L} = \begin{bmatrix} 0.278662 & 0.455625 & 0.549338 & 0.534216 \end{bmatrix}$$

$$\delta_{out} = \begin{bmatrix} -0.698582 & -0.952626 & 1.04574 & 1.03248 \end{bmatrix}$$

$$\delta_1 = \begin{bmatrix} -0.0675435 & -0.0921061 & 0 & 0.0998269 \\ -0.128683 & -0.175479 & 0 & 0.190189 \\ -0.0494353 & -0.0674128 & 0.0740024 & 0 \end{bmatrix}$$

$$\Delta w = \begin{bmatrix} -0.317837 & -0.649873 & 0.160052 \end{bmatrix}$$

$$\Delta W = \begin{bmatrix} -0.0491135 & -0.0935703 & -0.106061 \\ -0.0636314 & -0.12123 & 0.016283 \end{bmatrix}$$

$$w' = \begin{bmatrix} 0.12847 & 0.249193 & 0.05476 \end{bmatrix}$$

$$W' = \begin{bmatrix} 0.609787 & 0.717862 & 0.0385984 \\ 0.0224404 & 0.452736 & 0.876281 \end{bmatrix}$$

3 Exercise 3

3.1 Explain, in (your own) words, what this particular loss function is designed to achieve.

In a case that our classifier predicts some probable classes and assigns probabilities to them Hinge loss function will not punish this multi-prediction, as soon as, the prevailing probability is this of the correct class and the difference between the prevailing probability and the false class probability is within a chosen margin.

If we chose 1 for margin the every error is punished. If we chose 0 for margin then no error is being punished as soon as we predict the correct class.

This might be helpful for faster training and probably have our classifier over-fit less, since it reaches the local minimum faster.

3.2 Derive the gradient

$$\frac{\partial \bar{P}}{\partial \bar{O}} = \begin{bmatrix} \frac{\partial p_1}{\partial o_1} & \cdots & \frac{\partial p_1}{\partial o_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial p_N}{\partial o_1} & \cdots & \frac{\partial p_N}{\partial o_N} \end{bmatrix}$$

$$\mathcal{L}^{(1)}_{hinge} = \sum_{j \neq y} \max(0, p_j - p_{y_i} + margin)$$

$$p_i = \frac{exp(o_i)}{\sum_j exp(o_j)}$$

$$\frac{\partial p_i}{\partial o_j} = exp(o_i)\,' \cdot \frac{1}{\sum\limits_j exp(o_j)} + exp(o_i) \cdot (\frac{1}{\sum\limits_j exp(o_j)})\,' \tag{1}$$

In case i = j:

$$\frac{\partial p_i}{\partial o_j} = exp(o_i) \cdot \frac{1}{\sum\limits_j exp(o_j)} - exp(o_i) \cdot exp(o_j) \cdot \frac{1}{(\sum\limits_j exp(o_j))^2}$$

$$\frac{\partial p_i}{\partial o_j} = \frac{exp(o_i)}{\sum\limits_j exp(o_j)} - \frac{exp(o_i)}{\sum\limits_j exp(o_j)} \cdot \frac{exp(o_j)}{\sum\limits_j exp(o_j)}$$

$$\frac{\partial p_i}{\partial o_j} = p_i - p_i \cdot p_j$$

$$\frac{\partial p_i}{\partial o_j} = p_i \cdot (1 - p_j) \tag{2}$$

In case i $\neq$ j:

$$\frac{\partial p_i}{\partial o_j} = -exp(o_i) \cdot exp(o_j) \cdot \frac{1}{(\sum\limits_j exp(o_j))^2}$$

$$\frac{\partial p_i}{\partial o_j} = -p_i \cdot p_j \tag{3}$$

from (2) and (3):

$$\frac{\partial p_i}{\partial o_j} = \begin{cases} p_i \cdot (1 - p_j) & i = j \\ -p_i \cdot p_j & i \neq j \end{cases} \tag{4}$$

Using the Kronecker delta function

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

(4) becomes:

$$\frac{\partial p_i}{\partial o_j} = p_i \cdot (\delta_{ij} - p_j) \tag{5}$$

$$\frac{\partial \mathcal{L}_{hinge}}{\partial o_j} = \sum_{k \neq y_i} \max(0, p_k \cdot (\delta_{kj} - p_j) - p_{y_i} \cdot (\delta_{y_i j} - p_j) + margin)$$