

## Homework 4: Artificial Neural Networks

Lecturer: Georgios P. Karagiannis

georgios.karagiannis@durham.ac.uk

---

 Instructions: For Formative assessment, submit the solutions
 

---

**Exercise 1.** (★) Consider the multi-class classification problem, with a predictive rule  $h_w : \mathbb{R}^d \rightarrow \mathcal{P}$ , as a classification probability i.e.  $h_{w,k}(x) = \Pr(x \text{ belongs to class } k)$ , that receives values  $x \in \mathbb{R}^d$  returns vales in  $\mathcal{P} = \left\{ p \in (0, 1)^q : \sum_{j=1}^q p_j = 1 \right\}$ . Let  $h_w = (h_{w,1}, \dots, h_{w,q})^\top$ , let  $h_w(x)$  be modeled as an ANN

$$h_k(x) = \sigma_2 \left( \sum_{j=1}^c w_{2,k,j} \sigma_1 \left( \sum_{i=1}^d w_{1,j,i} x_i \right) \right)$$

for  $k = 1, \dots, q$ , and let the associated activation functions be

$$\sigma_2(a_k) = \frac{\exp(a_k)}{\sum_{k'=1}^q \exp(a_{k'})}, \text{ for } k = 1, \dots, q$$

(called softmax function) and  $\sigma_1(a) = \arctan(a)$ . Consider a loss

$$\ell(w, z = (x, y)) = - \sum_{k=1}^q y_k \log(h_{w,k}(x))$$

at  $w$  and example  $z = (x, y)$ , where  $x \in \mathbb{R}^d$  is the input vector (features), and  $y = (y_1, \dots, y_q)$  is the output vector (labels) with  $y \in \{0, 1\}^q$  and  $\sum_{k=1}^q y_k = 1$ . Consider that  $d$ ,  $c$ , and  $q$  are known integers.

**Hint:** You may use

$$\frac{d}{dx} \arctan(x) = \frac{1}{1+x^2}$$

- (1) Perform the forward pass of the back-propagation procedure to compute the activations which may be denoted as  $\{a_{t,i}\}$  and outputs which may be denoted as  $\{o_{t,i}\}$  at each layer  $t$ .
- (2) Show that

$$\frac{\partial}{\partial a_k} \sigma_2(a_j) = \sigma_2(a_j) (1(j=k) - \sigma_2(a_k))$$

$$\text{for } k = 1, \dots, q. \text{ Let } 1(j=k) = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}.$$

- (3) Perform the backward pass of the back-propagation procedure in order to compute the elements of the gradient  $\nabla_w \ell(w, (x, y))$ .

**Solution.**

- (1) Forward pass

**Set:**  $o_{0,i} = x_i$  for  $i = 1, \dots, d$

**Compute:**

**at**  $t = 1$ : for  $j = 1, \dots, c$

**comp:**  $\alpha_{1,j} = \sum_{i=1}^d w_{1,i,j} x_i$

**comp:**  $o_{1,j} = \arctan(\alpha_{1,j})$

**at**  $t = 2$ : for  $k = 1, \dots, q$

**comp:**  $\alpha_{2,k} = \sum_{j=1}^d w_{2,k,j} o_{1,j}$

**comp:**  $o_{2,k} = \frac{\exp(\alpha_{2,k})}{\sum_{k'=1}^q \exp(\alpha_{2,k'})}$

**get:**  $h_k = o_{2,k}$

(2) It is

$$\begin{aligned} \frac{d}{da_k} \sigma_2(a_j) &= \frac{d}{da_k} \frac{\exp(a_j)}{\sum_{j'} \exp(a_{j'})} = \begin{cases} \sigma_2(a_j) (1 - \sigma_2(a_j)) & j = k \\ -\sigma_2(a_j) \sigma_2(a_k) & j \neq k \end{cases} \\ &= \sigma_2(a_j) (1(j = k) - \sigma_2(a_k)) \end{aligned}$$

(3) It is

$$\frac{d}{da} \sigma_1(a) = \frac{1}{1 + a^2}$$

and

$$\begin{aligned} \frac{d}{da_k} \sigma_2(a_k) &= \sigma_2(a_j) (1(j = k) - \sigma_2(a_k)) \\ &= o_j (1(j = k) - o_k) \end{aligned}$$

and

$$\frac{d\ell_2}{do_{2,j}} = -y_j \frac{1}{o_{2,j}}$$

and

$$\begin{aligned} \frac{d\ell_2}{da_{2,k}} &= \sum_{j=1}^q \frac{d\ell_2}{do_{2,j}} \frac{do_{2,j}}{da_{2,k}} \\ &= \sum_{j=1}^q \left( -y_j \frac{1}{o_{2,j}} o_{2,j} (1(j = k) - o_{2,k}) \right) \\ &= \sum_{j=1}^q (-y_j (1(j = k) - o_{2,k})) \\ &= o_{2,k} - y_k \end{aligned}$$

**Backward pass:**

**at**  $t = 2$ : for  $k = 1, \dots, q$

**comp:**  $\tilde{\delta}_{2,k} = \frac{d}{d\alpha_{2,k}} \ell_T = o_{2,k} - y_k$

**at**  $t = 1$ : for  $j = 1, \dots, c$

**comp:**

$$\begin{aligned}\tilde{\delta}_{1,j} &= \frac{d}{d\xi} \sigma_1(\xi) \Big|_{\xi=\alpha_{1,j}} \sum_{k=1}^q w_{2,k,j} \tilde{\delta}_{2,k} \\ &= \left( \frac{1}{1 + \alpha_{1,j}^2} \right) \sum_{k=1}^q w_{2,k,j} \tilde{\delta}_{2,k}\end{aligned}$$

**Output:**

$$\frac{d}{dw_{1,j,i}} \ell = \tilde{\delta}_{1,j} x_i \text{ and } \frac{d}{dw_{2,k,j}} \ell = \tilde{\delta}_{2,k} o_{1,j}$$


---