

## Homework 4: Artificial Neural Networks

Lecturer: Georgios P. Karagiannis

georgios.karagiannis@durham.ac.uk

---

 Instructions: For Formative assessment, submit the solutions to all the parts of the Exercise.
 

---

**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.**