Instituto Superior de Agronomia, ULisboa
Green Data Science
Practical Machine Learning/Aprendizagem Automática Aplicada
Questionnaire #6, April 21, 2023

Name: ____________________

Topic: Overview of fundamentals of neural networks

Consider the following expressions:

$$g(z) = \frac{1}{1+e^{-z}} \tag{1}$$

$$\hat{y} = g\left(w_0 + \sum_{j=1}^{k} x_j\, w_j\right) \tag{2}$$

$$L_i(\mathbf{w}) = \mathcal{L}(y_i, \hat{y}_i), \text{ for } i = 1, \ldots, n \tag{3}$$

$$L(\mathbf{w}) = \frac{1}{B}\sum_{i=1}^{B} L_i(\mathbf{w}) \tag{4}$$

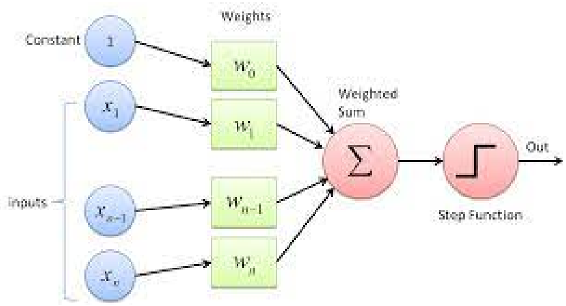
where y_i is the label ot the ith example. Each example is described by a vector $(x_1, \ldots, x_k)$, where each x is an explanatory variable. $\mathbf{w} = (w_0, w_1, \ldots, w_k)$ is a vector of weights. $\mathcal{L}$ is a function that takes two numbers and returns a measure of dissimilarity between them.

1. Equation 2 describes the a perceptron model. Is this a model for a regression problem or for a classification problem? How do you interpret the output?

Response: The output of the model is a value between 0 and 1 (since function g returns a value between 0 and 1) that can be interpreted as the probability of the example belonging to class "1". This model is therefore designed for a classification problem.

2. Draw a diagram that describes that perceptron model and label it with the symbols in Equation 2.

Response: The following diagram where instead of the step function we have g.



3. Indicate the symbols in the equations above that correspond to the following concepts, with a brief explanation,

a) Bias:

Response: w_0, is the constant term.

b) Batch size:

Response: B, is the nember of examples that are grouped in order to compute the loss.

c) Activation function:

Response: The activation function is represented by g. It is non linear.

d) Predicted label:

Response: The predicted label is the output the value returned by the model $\hat{y}$

e) Loss of a batch of examples:

Response: $L(\mathbf{w})$ is the loss computed for B examples.

4. What is the optimal set of weights for the perceptron model? Choose one option and give a brief explanation for the ones you did not choose.

a) The set of weights $\mathbf{w}^*$ that minimize $L(\mathbf{w})$

Response: correct answer

b) $\mathbf{w}^*$ such that the gradient $\nabla L(\mathbf{w}^*) = \dfrac{\partial L}{\partial \mathbf{w}}(\mathbf{w}^*)$ is zero

Response: A gradient converging to zero is an indication that the gradient descent algorithm could be approaching a local optimum but it does not guarantee that it is the best overall solution.

c) Given a current set of weights $\mathbf{w}^*$, and a learning rate η, it is the new set of weights defined by $\mathbf{w}^* - \eta\nabla L(\mathbf{w}^*)$

Response: This is just the updating rule of the gradient descent technique.