Artificial Intelligence Problems **(2017–2018)**
Third year – Grado en Ingeniería Informática,
Universidad de Sevilla

Artificial Neural Networks

Questions

Question 1. Answer if the following statements are TRUE or FALSE, in a reasoned way. If the answer is FALSE, give an example where the statement does not hold.

(a) Using as activation function the identity function, then a simple perceptron is able to learn the logical function AND.

(b) Using a simple perceptron with threshold as activation function, we can represent the function $g(A, B, C) \equiv ((A \vee B) \wedge \neg C)$.

Question 2. Design a simple perceptron with n input values and the threshold function as activation function, to calculate the MAJORITY-SIMPLE function; this function receives n inputs (that can be a 0 or a 1) and returns as output a 1 if there are strictly more 1s than 0s, or 0 otherwise.

Question 3. Justify geometrically that a simple perceptron is not enough to represent the XOR function. Give a multilayer neural network (with the threshold function as activation function) for that. Demostrar geométricamente que un perceptrón simple no puede calcular la función XOR. *Hint:* The XOR function can be obtained combining AND and OR, and these functions can be represented using perceptrons.

Question 4. The function PARITY receives n bits as input and returns 1 if there are an even number of 1s, 0 otherwise. Is it possible to represent the PARITY function using a simple perceptron? And using a neural network with one hidden layer (with threshold activation function)?
Hint: Use a hidden layer with n units, where each unit i activates only if if it receives as input more than i 1s.

Question 5. Let f be a function from $R \times R$ to $\{-1, 1\}$ such that $f(-1, 1) = f(0, 0) = 1$ and $f(-1, 0) = f(0, 1) = -1$. Suppose that using those examples as training set, we apply the perceptron training rule (with bipolar activation function). Will the algorithm be able to find adequate weigths for the perceptron to compute that function? What about if we use the delta rule with sigmoid activation function?

Question 6. Answer if the following statements are TRUE or FALSE, in a reasoned way.

(a) If the training set is linearly separable, then the perceptron learning algorithm, with threshold activation function, converge to weights that give us the minimum of the squared error

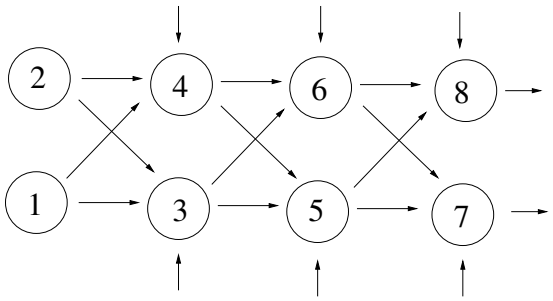
(b) If we apply one step of the delta rule algorithm, then it is guaranteed that the global squared error decreases

(c) If we apply one step of the gradient descent algorithm, then it is guaranteed that the global squared error decreases

Problems

Ejercicio 1. Consider a perceptron with weights $w_0 = w_1 = w_2 = 0{,}5$, and sigmoid activation function. Let D be the training set $D = \{\langle(1,0),1\rangle, \langle(0,1),0{,}5\rangle\}$. Compute the squared error variation after one step the gradient descent algorithm (with learning rate $\eta = 0{,}8$).

Ejercicio 2. Consider the following artificial neural network:



Assume that the initial weights are 1.0, that we use the sigmoid activation function, and that the learning rate is $\eta = 0{,}1$. Compute the weight w_{06} after one step of the backpropagation algorithm corresponding to the training example $\langle(1,1),(1,1)\rangle$

Ejercicio 3. Consider a perceptron with weights $\vec{w} = (w_0, w_1, w_2)$, $w_0 = w_1 = w_2 = 0{,}8$ and sigmoid activation function. If we have a training set $D = \{E_1, E_2\}$ con $E_1 = \langle(0,1),1\rangle$ and $E_2 = \langle(1,0),0\rangle$.

(a) Compute the squared error $E(\vec{w})$, with respect to D.

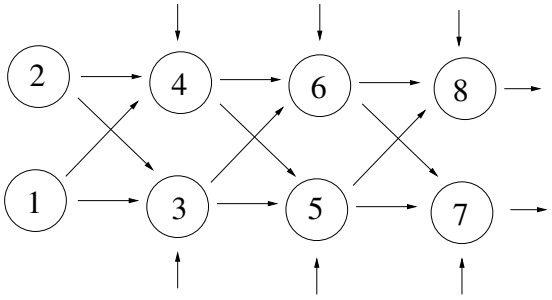
(b) Compute the weight w_2 after *one update* carried out by the GRADIENT DESCENT algorithm, with a learning rate $\eta = 0{,}2$.

Ejercicio 4. Apply the GRADIENT DESCENT algorithm (with $\eta = 0{,}1$) *until the first update of the weight* w_2, using the training set:

	x_1	x_2	x_3	y
e_1	0,7	0,2	0,1	0,3
e_2	0,3	0,5	0,2	0,8
e_3	0,1	0,1	0,8	0,6

Consider initial weights $w_0 = 0{,}1$, $w_1 = 0{,}1$, $w_2 = 0{,}1$ and $w_3 = 0{,}1$.

Ejercicio 5. Consider the following neural network that uses the sigmoid as activation function.



Assume that initially all the weights are 1.0, and that we use a learning factor $\eta = 0{,}1$. Compute the value $w_{3,6}$ after one step of the backpropagation algorithm, corresponding to the example $\langle(0{,}5, 0{,}5), (0{,}8, 0{,}8)\rangle$?

Ejercicio 6.

Let f be a function from $R \times R$ to $\{-1, 1\}$. Suppose that we want to learn f by a perceptron with bipolar activation function, and that we have the following training set:

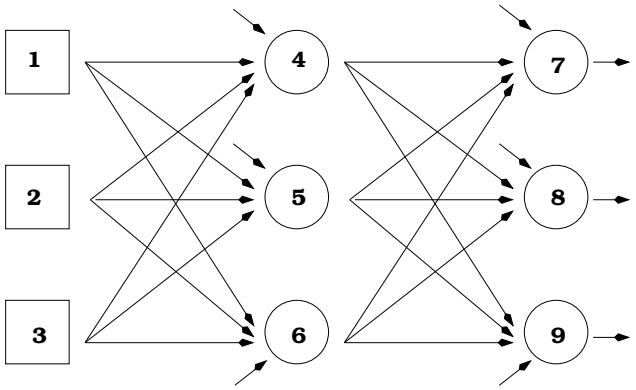
	Input	Output
E_1	$(2, 0)$	1
E_2	$(0, 0)$	-1
E_3	$(2, 2)$	1
E_4	$(0, 1)$	-1
E_5	$(1, 1)$	1
E_6	$(1, 2)$	-1

Apply the perceptron learning algorithm using that training set, taking the examples in the order they appear in the table. Take 0 as the initial value for all the weights and learning factor 0,1.

Using the learned perceptron, compute the output for the following inputs: $(0, 2)$, $(1, 0)$ and $(2, 1)$?

Ejercicio 7.

Consider the following neural network, that uses the sigmoid as activation function:



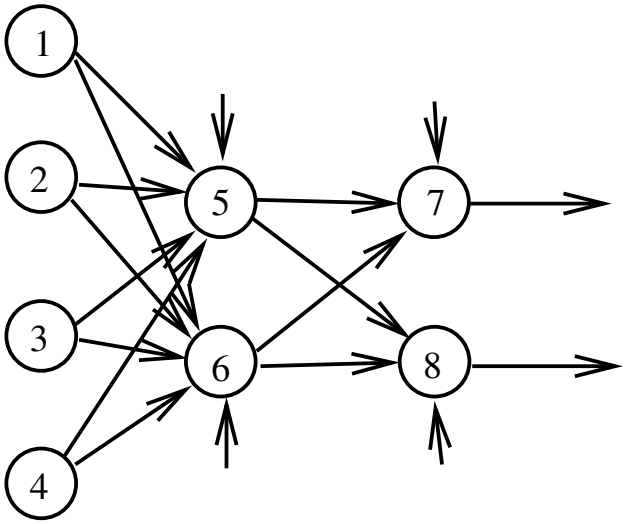
1. Compute the output for the input $\vec{x} = \langle 1, 0, 0 \rangle$, assuming that all the weights are 0,5.

2. Consider now another example $\vec{x} = \langle x_1, x_2, x_3 \rangle$ with expected output $\vec{y} = \langle 1, 0, 0 \rangle$. Assume that in all the units of the hidden layers, the output is $a_i = 0{,}2$, and that the output in all the units of the output layer is $a_i = 0{,}5$. With these assumed values, compute the error Δ_4 for that example (as computed by the backpropagation algorithm). Give also the formulas for updating the weights $w_{0,4}$, $w_{4,7}$, $w_{4,8}$ and $w_{4,9}$ (without computing them).

3. Assume that this network models a classiffier for vectors $\vec{x} \in \mathbb{R}^3$,and that we have already learned the weights using a training set.

   - Explain how we classify a new example $\vec{x} = \langle x_1, x_2, x_3 \rangle$.
   - If we were interested in classifying the examples in only two classes, would you have used a different network structure?

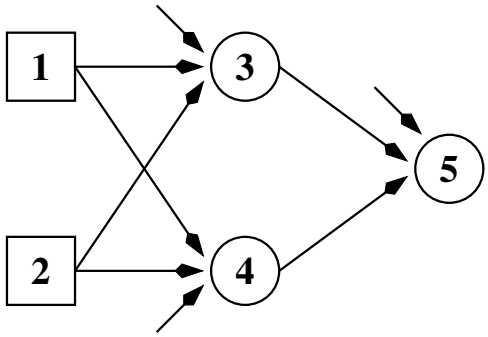
Ejercicio 8. Consider the following neural network that uses the sigmoid as activation function:



Suppose we have and example (x_1, x_2, x_3, x_4) with expected output (y_7, y_8). Assume that we have already computed the corresponding output a_i for each unit, $i = 1, \ldots, 8$. Give the formulas applied by the backpropagation algorithm to compute the errors Δ_8, Δ_7 y Δ_6, and the formulas to update the weights $w_{6,7}$ and $w_{6,8}$.

Ejercicio 9.

Consider tyhe following neural network, with activation function $g(x) = x$:



If $(\vec{x_1}, y_1) = (\langle 0, 0\rangle, 1)$ and $(\vec{x_2}, y_2) = (\langle 1, 1\rangle, 0)$ are two examples of the training set, apply the backpropagation algorithm for these two examples, with learning rate 0,1, and assuming that all the initial weights are 0,5.

Ejercicio 10.

Explain how a neural network could be used for recognition of handwritten letters. Describe precisely which network structure you would use, how would you represent both the input and output of the network, how would you obtain the trainig set and how would you train the network to obtain suitable weights. What do we understand in this context by "suitable weights"?

Ejercicio 11.

An *electronic nose* analyzes, using sensors, the vapors originating from certain substances and classifies them on the basis of the quantitative information obtained. Suppose we use 16 sensors for identify four types of red wines: Cabernet, Merlot, Syraz and Tempranillo.

Describe in detail how we could use a neural network to address this problem: how could we obtain a training set and how would you encode the examples, the network structure, the training algorithm used, and how the network could be used once trained, to identify a new red wine.

Ejercicio 12.

Suppose that a cable television company wants to design an automated system to recommend one of its five channels to its customers, depending on their preferences, which are to be guessed from a survey with 20 questions. How would you design the system using a neural network? What structure would this neural network have? How do you train the network and what would your training set be? Once trained, how would you use the network obtained to recommend a thematic channel to a new customer?

Ejercicio 13.

Suppose we want to design a system of "customized ads". for users of a web portal. In this website there are twenty topic sections and the company manages four possible advertising profiles, which are supposed to depend on the degree of interest each user have for each topic. Describe how a neural network could be used to implement this system: structure of the network, training set, learning the weights, use of the network once trained...

Ejercicio 14.

Suppose we want to design an automated system to recognize the people's moods watching their faces. By simplifying things, let's say we consider four different types of moods: cheerful, sad, angry and neutral. Assuming that our system has a camera that is capable of obtaining digitized images of a person's face, how would you design the system using a neural network?

Describe precisely which network structure you would use, how would you represent both the input and output of the network, how would you obtain the trainig set, how would you compute suitable weights and how would you use the network once klearned.

Ejercicio 15.

Suppose that we want to learn a classifier for vectors in R^n, and that we have two possible classes. We do not know an exact "formula" for doing that classification, but we know the classification of some vectors (that is, a training set).

- Give a real example of the situation described above
- If the training set were not linearly separable and the classifier were a perceptron, hoe many input and output units would you use? which activation function would you use? Once learned, how would you use it to classify new examples?
- Which learning algorithm would you use?
- For $n = 3$, perform one step of that learning algorithm *considering one positive example* (use an invented example, initial weights and learning factor).

Now assume that we have **four possible classes**

- Give a real example of that situation.

- If we wanted to learn a multilayer network, how many units you will need in the input layer and how many in the output layer? Which activation function would you nuse? Once learned, how would you use the network to classify new instances?
- Consider specifically $n = 3$ and a multilayer network with two units in the hidden layer. In that situation, and using the backpropagation algorithm to learn the weights, describe (symbolically) the formulas used by the backpropagation algorithm to update the weights in every step.