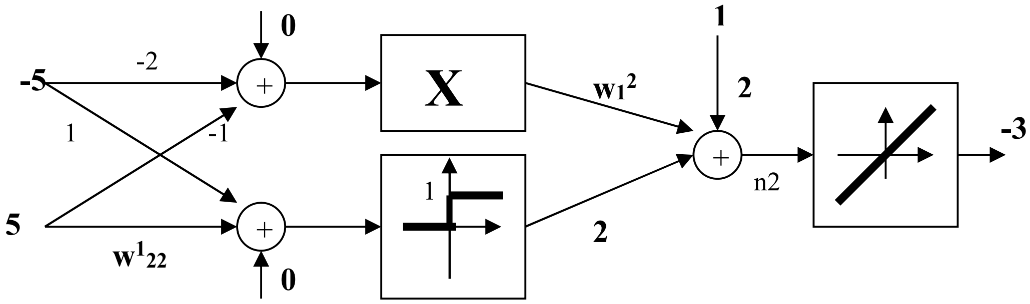
DEI/FCTUC
Computational Learning/Neuronal and Fuzzy Computation

Duration: **120 minutes**

Exam with consultation, except past exams and its solutions. It is not allowed the use of computers.

(Remark: this document is intended to support the study for the exam. It does not mean that the exam questions will be similar to the ones in this document. If you find mistakes, please inform me AD).

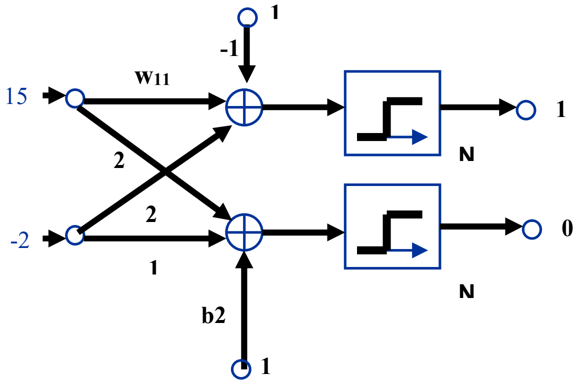
1. Consider the neural network of the figure with arbitrary weights w^1_{22} and w^2_1 to be computed in each case in order to obtain the desired output.



a) Which of the following activation functions are possible in the neuron **X** ?

Function	Yes or No	If yes, compute a set of values for these weights. If not, justify.
Logsig	Yes	There are many possible solutions. Start by the output, then n2=-3. There are many ways to obtain this -3 from the sum of the two neurons in the first layer plus the bias=2. For example, one solution is: $w_{22}^1=2$, $w_1^2=-7(1+e^{-5})$. Note that in the answer in the exam these values of the weights must be proved.
Hardlims (symmetric)	Yes	There are also many possible solutions. For example $w_{22}^1=2$, $w_1^2=-7$ Note that in the answer in the exam these values of the weights must be proved.
Purlin	Yes	Again, there are many solutions. Example $w_{22}^1=2$, $w_1^2 = -7/5$ Note that in the answer in the exam these values of the weights must be proved.

2. Consider the following perceptron. Compute the interval of values of the weight w_{11} and of the bias b_2 that allow to obtain the classification shown in the figure.

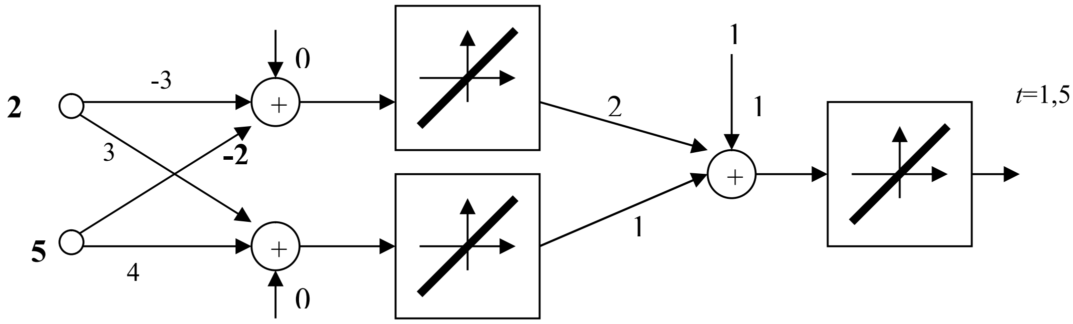


We will have

$15w_{11}$-2*2-1*1>=0 , from where w_{11}>=5/15, or w_{11} in [5/15, ∞ [

15*2-2*1+b2<0, from where b2<-28 , or, b2 in]-∞, -28[

3. a) Make an iteration of the retropropagation algorithm to adjust the weight w^1_{12}, starting from its initial value underscored in the figure. The target for the input is 1,5.



$F = e^2$ (objective function, criteria) $\quad a = n^2$

$$\mathrm{w}_{12}^1(k+1) = \mathrm{w}_{12}^1(k) - \alpha \frac{\partial F}{\partial w_{12}}\Big|_k \qquad n^2 = a_1^1 \times 2 + a_2^1 \times 1 + 1$$

$$n_1^1 = 2 \times (-3) + 5 w_{12}^1 + 0$$

$$a_1^1 = n_1^1$$

1st step, forward

$n_1^1 = -3 \times 2 - 2 \times 5 + 0 = -16 \qquad a_1^1 = -16 \qquad n^2 = -2 \times 16 + 1 \times 26 + 1 = -5$

$n_2^1 = 3 \times 2 + 4 \times 5 + 0 = 26 \qquad a_2^1 = 26$

2nd step, calculation of the error: $e = t - a = 1{,}5 - (-5) = 6{,}5$

3rd step: retropropagation of the error and updating of the weight

$$\frac{\partial F}{\partial w_{12}^1} = \frac{\partial F}{\partial e} \times \frac{\partial e}{\partial a^2} \times \frac{\partial a^2}{\partial n^2} \times \frac{\partial n^2}{\partial a_1^1} \times \frac{\partial a_1^1}{\partial n_1^1} \times \frac{\partial n_1^1}{\partial w_{12}^1}$$

$$= 2e \times (-1) \times \ 1 \ \times \ 2 \ \times \ 1 \ \times \ 5$$

$$\frac{\partial F}{\partial w_{12}^1} = -20e = -130$$

$$w_{12}^1(1) = w_{12}^1(0) - \alpha \frac{\partial F}{\partial w_{12}^1}\Big|_0 = -2-(-130) = 128.$$

(for a learning coefficient equal to 1).

b) Discuss the strategy to chose the learning coefficient.

This learning coefficient determines the convergence of the retropropagation algorithm. If it is too small, the convergence is very slow for a local minimum if eventually reaches one). From iteration to iteration the weights are changing slowly. If it is too big, the values of the weights vary considerably from iteration to iteration, as is the case in the example where the weight varies from -2 to 128 I the first iteration. Probably if the following iterations, if nothing is done, this value may oscillate from one side to the other, without converging (or diverge, or converge so slowly that it would not have practical utility).

At the start one can use an appreciable value (not too low. There is one guideline: its value should be less than the inverse of the maximum eigenvalue of the input matrix. Then, as the minimum is approached, it should be reduced, for example dividing by 2 successively, until no significative improvement of the criteria is obtained. In this procedure we would have an adaptive learning coefficient.

In a), if we proceed for the second iteration with the same learning coefficient equal to 1, the error will be 1293.5, much bigger than the initial, meaning that we are in a divergent trajectory. The learning coefficient must then be substantially reduced, for example to 0.1.

4. Consider four objects in the bidimensional space represented by the following four points given by their coordinates O1 (3,4), O2 (-1,4), O3 (1,3), O4 (-1,1) - entre. Project and design a perceptron to classify these objects into four classes, using a minimum number of neuron. Design the complete perceptron.

Use the delivered millimetric paper.

5. Consider the fuzzy sets

$$\underset{\sim}{A}=\left\{\frac{0,6}{2}+\frac{0,7}{5}+\frac{1}{8}\right\} \qquad \underset{\sim}{B}=\left\{\frac{0,9}{M}+\frac{0,7}{N}+\frac{0,5}{P}\right\}$$

Defined in the respective universes of discourse.

a) Consider the Zadeh implication, $\underset{\sim}{A} \Rightarrow \underset{\sim}{B}$. Compute the respective implication matrix.

b) Using *modus ponens* compute the value of the consequent of the implication for the new antecedent $\underset{\sim}{A}' =\left\{\frac{0,2}{2}+\frac{0,5}{5}+\frac{0,8}{8}\right\}$ (use *max-prod* composition, if needed, and the implication matrix computed in a).

$$\underset{\sim}{A}=\left\{\frac{0,6}{2}+\frac{0,7}{5}+\frac{1}{8}\right\} \qquad \underset{\sim}{B}=\left\{\frac{0,9}{M}+\frac{0,7}{N}+\frac{0,5}{P}\right\}$$

Relational matrix relacional (implication of Zadeh): $\mu_{\underset{\sim}{R}}(x,y)=\max[\min(\mu_{\underset{\sim}{A}}(x),\mu_{\underset{\sim}{B}}(y)),1-\mu_{\underset{\sim}{A}}(x)]$

$$\underset{\sim}{R}=\begin{bmatrix} 0,6 & 0,6 & 0,5 \\ 0,7 & 0,7 & 0,5 \\ 0,9 & 0,7 & 0,5 \end{bmatrix}$$

Composition

$$\underset{\sim}{A}' \circ \underset{\sim}{R}=\begin{bmatrix} 0,2 & 0,5 & 0,8 \end{bmatrix} \circ \begin{bmatrix} 0,6 & 0,6 & 0,5 \\ 0,7 & 0,7 & 0,5 \\ 0,9 & 0,7 & 0,5 \end{bmatrix} = \begin{bmatrix} 0,72 & 0,56 & 0,4 \end{bmatrix}$$

$$\underset{\sim}{B}' =\left\{\frac{0,72}{M}+\frac{0,56}{N}+\frac{0,4}{P}\right\}$$

6. A system has two inputs P1 e P2 and one output S1, with the following scales:

P1: [0 5] P2: [-5 0] S: [-10 10]

Define in the normalized interval [-1 1] three triangular fuzzy sets, with an adequate coverage property, for each of the variables (inputs and output). Label them as Negative (N) , Zero (Z), Positive (P).

Consider the following Mamdani rules:

Rule 1: IF P1 is Negative AND P2 is Zero THEN S1 is Zero ($N \wedge Z \rightarrow Z$)
Rule 2: IF P1 is Positive AND P2 is Positive THEN S1 is Positive ($P \wedge P \rightarrow P$)
Rule 3: IF P1 is Zero and P2 is Positive THEN S1 is Positive ($Z \wedge P \rightarrow P$)

Consider now the measured inputs

Input P1= 0,75 Input P2 = -2,0

Compute the fuzzy output of each rule and the total fuzzy output, using the minimum for logic AND, the product for the inference and the defuzification by the height method.

Answer:

Rule 1 ($N \wedge Z \rightarrow Z$)

For P1=0,75 (-0,7 normalized), the membership to N is 0.7
For P2=-2 (0,2 normalized), the membership to Z is 0.8 minimum =0.7
The output of the rule is a triangle with height 0.7 resulting from the product of Z by 0.7.

Rule 2 ($P \wedge P \rightarrow P$)

For P1=0,75 (-0,7 normalized), membership to a P is 0
For P2=-2 (0,2 normalized), membership to P é 0.3 minimum = 0
The output of the rule is a triangle with height 0.

Rule 3 ($Z \wedge P \rightarrow P$)

For P1=0,75 (-0,7 normalized), membership to Z is 0.3
For P2=-2 (0,2 normalized), membership to P é 0.2 minimum =0.2
The output of the rule is a triangle of height 0.2 resulting from the product of P by 0.2.

The total output are two triangles resulting from the rule 1 and rule 3. If we apply the defuzification by the height, it becomes zero, normalized, or 0 denormalized; i.e, the total output is 0.

7. Consider now the ANFIS architecture, TSK of order zero. Design the network needed to implement the equivalent of the fuzzy system of the previous question considering the consequents as the centers of the membership functions of the consequents of question 6 and using the product as the conjunction operator.
Normalize previously the inputs and, if needed, the denormalization of the output at the end. Note in the figure the computations that the ANFIS network makes.
Compare with the results of question 6.

The consequents of the rules are 1 and 0, i-is, we aill have:

Rule 1: IF P1 is Negative AND P2 is Zero THEN S1 is 0 $(N \wedge Z \rightarrow Z)$
Rule 2: IF P1 is Positive AND P2 is Positive THEN S1 is 1 $(P \wedge P \rightarrow P)$
Rule 3: IF P1 is Zero AND P2 is Positive THEN S1 is 1 $(Z \wedge P \rightarrow P)$

Normalized inputs: P1=-0,7 P2=0,2.
Layer 1: fuzzification: P1 é 0.7 N and 0.3 Z e 0 P
P2 é 0.8 Z e 0.2 P e 0 N

Layer 2: rule 1 : 0.7*0.8=0.56
rule 2: 0*0.3=0
rule 3: 0.3*0.2*=0.06

Layer 3: firing normalization
Rule 1 gives a1=0.9
Rule 2 gives a2=0
Rule 3 gives a3= 0.1

Layer 4: calculation of the normalized output
Rule 1: y1=0, a1*y1=0
Rule 2: y2=1, a2 *y2=0
Rule 3: y3=1, a3*y3=0.1

Layer 5: sum of the ouputs

ytotal=0.1 normalized

Denormalized output: 0.1*10=1

The output is different from the one in 6; this is expectable since TSK systems do not work as Mamdani ones and the conjunction operator here is different, the product instead of the minimum as in 6. In the case the defuzification method used in 6 (height) introduces additional error, since the small positive triangle in 7 has no influence on the result. If it would have, then the output in 6 would be a bit positive. For example if in 6 we make ytotal the weighted average of the picks of the triangles S_1 and S_3

$ytotal = \frac{0.7 \times 0 + 0.2 \times 1}{0.7 + 0.2} = 0.22$ also different from the one in 7.

Note that if in 8 the conjunction of the antecedents would be the minimum, then in 7 ytotal=0.22.