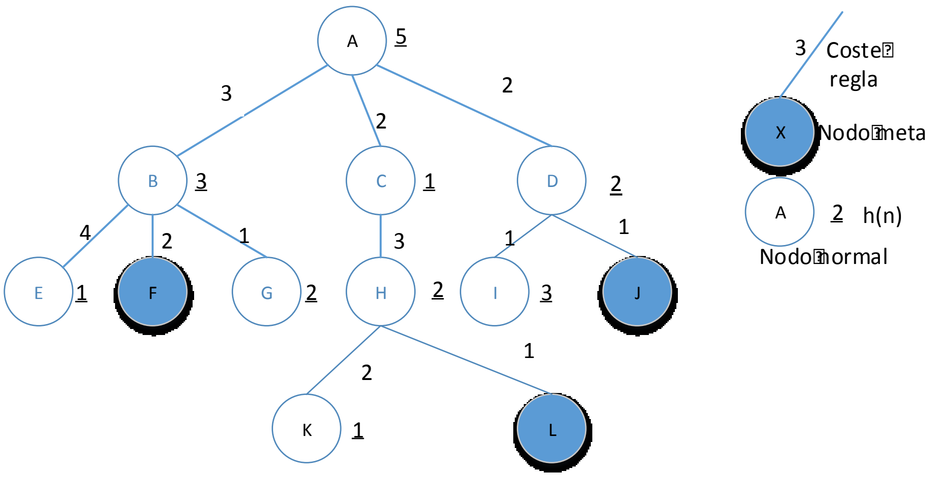
Intelligent Systems – Final Test (Block 1)
ETSINF, Universitat Politècnica de València
January 26, 2017 (2 points)

Surnames:							
Name:							
Group:	A	B	C	D	E	F	FLIP

1) Given the search space of the figure, if we apply a search of type A ($f(n)=g(n)+h(n)$) how many nodes are generated to find the solution?



A. 7
B. 8
C. 10
D. 12

2) If we apply a A* search in CLIPS, rules must not have a `retract` command in the RHS because:

A. Facts are retracted and so we would not be able to compute the value of $g(n)$
B. We would not be able to explore paths other than the path selected at first place
C. We would not be able to compute the optimal solution
D. None of the above

3) Given a search algorithm of type A, ($f(n)=g(n)+h(n)$), indicate the **CORRECT** statement:

A. If $h(n)$ is consistent (and admissible), the algorithm will always expand fewer nodes than an uninformed search
B. If $h(n)$ is consistent (and admissible), the algorithm will always expand fewer nodes than when $h(n)$ is not consistent
C. Whether or not $h(n)$ is admissible, the same solution will always be found
D. None of the above

4) Given a RBS composed of the following rule:

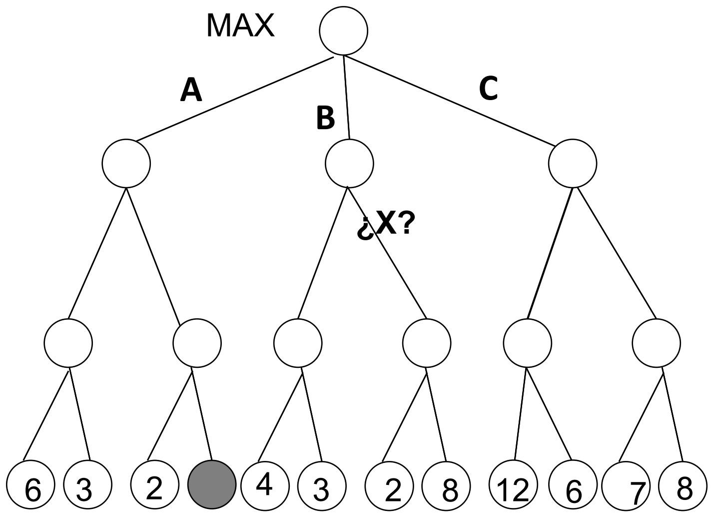
```
(defrule rule-1
   ?f <- (lista ?y $?x ?y $?x ?y)
=>
   (retract ?f)
   (assert (lista $?x)))
```

, and the initial Working Memory {(lista 1 2 3 2 3 2 3 2 3 2 1 2 3 2 3 2 3 2 3 2 1)}, which is the final state of the WM?

A. {(lista 3 2 3)}
B. {(lista 1) (lista)}
C. {(lista 2 1 2)}
D. {(lista 1)}

---

5) Given the game tree of the figure and assuming we apply an alpha-beta algorithm:



which is the value that the shadowed node should have in order to get the cut-off of the figure?

A. Any value
B. A value lower than 3
C. A value higher or equal than 4
D. The cut-off would never be produced (or none of the above answers)

---

6) Given the game tree of the above figure and assuming the cut-off is produced, after applying an alpha-beta algorithm:

A. MAX will choose branch A
B. MAX will choose branch B
C. MAX will choose branch C
D. MAX will choose either branch A or B

Intelligent Systems – Problem Block 1
ETSINF, Universitat Politècnica de València,
January 26, 2017 (3 points)

We want to design a RBS for handling the collection of stickers of a number of children. A sticker is represented by an alpha-numerical identifier (e.g., A1, B3, C2, etc.). A child has a number of stickers (including repeated stickers) and the number of children is not limited in the application. An example is:

- Kid 1 has the stickers: A2 A4 A5 B1 A2 B3
- Kid 2 has the stickers: B3 A4 C2 C1 B3 C2
- Kid 3 has the stickers: C2 C4 B1 A2

The dynamic information of the problem is represented with facts that follow this pattern:

(collections [kid ?n [?id-sticker]^m fkid ?n]^m)

where:

?n ∈ INTEGER ;; kid identifier
?id-sticker ∈ {A1, A2, B1,…} ;;sticker identifier

Answer the following questions:

a) (0.3 points) Write the facts of the Working Memory that represent the above example.

```
(deffacts datos
    (collections  kid 1 A2 A4 A5 B1 A2 B3 fkid 1
             kid 2 B3 A4 C2 C1 B3 C2 fkid 2
             kid 3 C2 C4 B1 A2 fkid 3))
```

b) (1 point) Write a single rule for two kids to swap a sticker. Swapping a sticker is only allowed if kids hand over a repeated sticker of their collections and they get in return a sticker which is not in their collections.

```
(defrule intercambio
    (collections $?x kid ?n1 $?c1 ?c $?c2 ?c $?c3 fkid ?n1
             $?y kid ?n2 $?p1 ?p $?p2 ?p $?p3 fkid ?n2 $?z)
    (test (neq ?c ?p))
    (test (and (not (member$ ?c $?p1))(not (member$ ?c $?p2))(not (member$ ?c $?p3))))
    (test (and (not (member$ ?p $?c1))(not (member$ ?p $?c2))(not (member$ ?p $?c3))))
    =>
    (assert (collections $?x kid ?n1 $?c1 ?c $?c2 ?p $?c3 fkid ?n1 $?y kid ?n2 $?p1 ?p $?p2 ?c
$?p3 fkid ?n2 $?z)))
```

Another solution:

```
(defrule intercambio
  ?f1 <- (collections $?x kid ?n1 $?c1 ?c $?c2 ?c $?c3 fkid ?n1
                 $?y kid ?n2 $?p1 ?p $?p2 ?p $?p3 fkid ?n2 $?z)
  ?f2 <- (collections $?x kid ?n1 $?todo1 fkid ?n1
                 $?y kid ?n2 $?todo2 fkid ?n2 $?z)
  (test (eq ?f1 ?f2))
  (test (neq ?c ?p))
  (test (and (not (member$ ?p $?todo1))(not (member$ ?c $?todo2))))
 =>
  (assert (collections $?x kid ?n1 $?c1 ?c $?c2 ?p $?c3 fkid ?n1 $?y kid ?n2 $?p1 ?p $?p2 ?c
$?p3 fkid ?n2 $?z)))
```

c) (0.7 points) Write a single rule to display the children that have a sticker exactly three times in their respective collections. The rule must display a message per kid and sticker; example: "The kid " ?n " has the sticker " ?x " three times".

```
(defrule acaparador
  (collections $? kid ?n $?x1 ?y1 $?x2 ?y1 $?x3 ?y1 $?x4 fkid ?n $?)
  (test (and (not (member ?y1 $?x1))(not (member?y1 $?x2))
             (not (member ?y1 $?x3))(not (member ?y1 $?x4))))
 =>
 (printout t "The kid " ?n " has the sticker " ?y1 "  three times " crlf))
```

d) (1 point) Let's suppose that the WM contains facts that follow the pattern (special ?id-sticker) to indicate that the sticker identified with ?id-sticker is a special sticker. Write a single rule to calculate the number of children who have at least two special and different stickers. The result of executing the rule will be a fact that follows this format: (special-list [?n]^{m}), where ?n is the identifier of a kid who has at least two different special stickers. The kid identifier must appear only once in the list (even though the kid has several special stickers). Assume that the WM contains several facts with the format (special ?id-sticker) and the fact (special-list).

```
   (defrule especiales
      (collections $? kid ?n1 $? ?a $? ?b $? fkid ?n1 $?y)
      (special ?a)
      (special ?b)
      (test (neq ?a ?b))
  ?r <- (special-list $?z)
      (test (not (member$ ?n1 $?z)))
   =>
     (retract ?r)
     (assert (special-list $?z ?n1)))
```

Intelligent Systems: Final Exam Block 2

ETSINF, Universitat Politècnica de València, January 26th, 2017

Surname(s): ☐ **Name:** ☐

Group: ☐ **3A** ☐ **3B** ☐ **3C** ☐ **3D** ☐ **3E** ☐ **3F** ☐ **FLIP**

Tick only one choice among the given options (2 points, estimated time: 30 minutes).

1 $\boxed{\text{C}}$ Let X and Y be two random variables, and let $P(X,Y)$, $P(X \mid Y)$, $P(Y \mid X)$, $P(X)$ and $P(Y)$ be joint, conditional and unconditional probabilities of X and Y. Indicate which of the following statements **IS NOT CORRECT**.

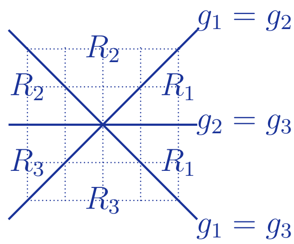
A) Both, $P(X)$ and $P(Y)$, can be obtained from $P(X,Y)$.

B) Both, $P(X \mid Y)$ and $P(Y \mid X)$, can be obtained from $P(X,Y)$.

C) $P(Y \mid X)$ can be obtained from $P(X \mid Y)$ and $P(X)$ without knowing $P(Y)$.

D) $P(Y \mid X)$ can be obtained from $P(X \mid Y)$ and $P(Y)$ without knowing $P(X)$.

2 $\boxed{\text{C}}$ For a three-class problem in $\mathbb{R}^2$, we have a classifier defined by three discriminant linear functions: $g_1(\mathbf{x}) = x_1$, $g_2(\mathbf{x}) = x_2$ and $g_3(\mathbf{x}) = -x_2$. Show the expression that **IS NOT CORRECT** for the defined classifier.



A) It defines three decision boundaries that intersect in the origin coordinate $(0,0)$.

B) The decision region of class 1 is defined as $R_1 = \{\mathbf{x} \in \mathbb{R}^2 : x_1 > 0 \wedge x_1 > |x_2|\}$

C) In the decision region R_2, x_2 is lower than 0 and in R_3, x_2 is greater than 0.

D) In the decision region R_2, x_2 is greater than 0 and in R_3, x_2 is lower than 0.

3 $\boxed{\text{B}}$ Let $\hat{p}$ be the probability of error of a classifier estimated on a test set of size N and let $I = [\hat{p} \pm \epsilon]$ be the confidence interval of this estimation. Indicate the **CORRECT** answer.

A) If $N = 160$ and the classifier makes at least one error, ϵ will be less than 1 %.

B) If $N > 150$ and the probability of error is $\hat{p} = 0.1$, ϵ will be less than 5 %.

C) If N_e is the number of errors made by the classifier, then $\hat{p} = N/N_e$ and ϵ is inversely proportional to N.

D) It is not possible to determine ϵ if $\hat{p} = 0$.

4 $\boxed{\text{D}}$ We applied the K-means algorithm on a set of two-dimensional objects to obtain a partition into two clusters. After a number of iterations of the K-means algorithm, the following partition into two clusters was obtained: : $\{\{(0,1)^{\mathbf{t}}, (0,2)^{\mathbf{t}}\}, \{(0,3)^{\mathbf{t}}, (0,5)^{\mathbf{t}}, (0,6)^{\mathbf{t}}, (0,7)^{\mathbf{t}}, (1,6)^{\mathbf{t}}, (-1,6)^{\mathbf{t}}\}\}$. Show the **CORRECT** statement:

A) The Sum of Square Errors (SSE) is 15 and it can reach a value of 8

B) The value of SSE is 15 and it will be 12 after convergence of the K-means algorithm

C) The value of SSE is 12 and it will be 10 after convergence of the K-means algorithm

D) The value of SSE is 12 and it will be 6 after convergence of the K-means algorithm

5 $\boxed{\text{D}}$ Let M be a hidden Markov model and let x be a string of length T to which M assigns a positive probability. Let q be a regular (non-final) state of M and let t be a point in time not greater than T. Consider the $\alpha(q,t)$ value computed by the *Forward* algorithm and the $V(q,t)$ computed by the Viterbi algorithm. Suppose that both $\alpha(q,t)$ and $V(q,t)$ are positive. Show the **CORRECT** answer in connection with $\alpha(q,t)$ and $V(q,t)$:

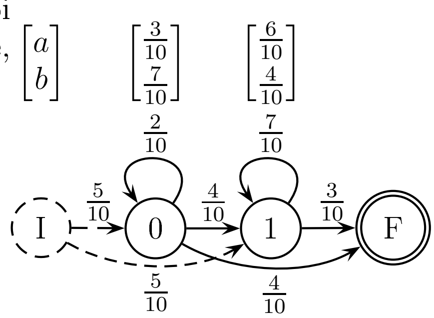
A) Both values will always be the same if $t > 1$.

B) Both values will never be the same if $t > 1$.

C) Both values will never be the same if $t = 1$.

D) Both values will always be the same if $t = 1$.

6 $\boxed{\text{A}}$ Given the Markov model M of the figure, the application of the Viterbi algorithm to the string "`bab`" returns an approximate probability value, $\tilde{P}_M(bab)$, such that:



A) $0.000 \leq \tilde{P}_M(bab) < 0.010$

B) $0.010 \leq \tilde{P}_M(bab) < 0.015$

C) $0.015 \leq \tilde{P}_M(bab) < 0.020$

D) $0.020 \leq \tilde{P}_M(bab)$

$$\tilde{P}_M(bab, 011F) = \tilde{P}_M(bab, 111F) = \frac{70560}{10^7} = 0.007056$$

Intelligent Systems: Final Exam Block 2

ETSINF, Universitat Politècnica de València, January 20th, 2016

Problems (3 points; estimated time: 45 minutes)

Let us consider a classification problem of two classes, 0 and 1, for objects represented in $\{0,1\}^2$, that is, bit vectors defined as $\mathbf{x} = (x_1, x_2)^t$ with $x_1, x_2 \in \{0,1\}$. In addition, there are four training samples available:

$\mathbf{x}_n$	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
x_{n1}	0	0	1	1
x_{n2}	0	1	0	1
c_n	0	1	1	0

To do:

1 (0.75 points) Show a trace of the Perceptron algorithm running for one iteration with initial weight vectors equal to zero, learning rate $\alpha = 1$ and margin $b = 0.1$. What are the weight vectors obtained at the end of the iteration?

2 (0.50 points) From the initialization given above, will the Perceptron algorithm converge to a solution without misclassified training samples?
Please say yes or no and then briefly discuss the answer.

3 (0.25 points) Is there any initialization with non-null weight vectors, $\alpha > 0$ and $b = 0.1$, from which the Perceptron algorithm will converge to a solution without misclassified training samples?
Please say yes or no and then briefly discuss the answer.

4 (0.75 points) Apply the DCT learning algorithm shown in class to the classification problem given. In order to measure the node impurity, use the entropy of the empirical distribution of the class posterior probabilities at the node. On the other hand, to decide whether a node is terminal or not, use the stop criterion discussed in class with impurity (decrement) threshold of $\epsilon = 0.1$. Also, to explore possible partitions ("splits") of a node, consider only a threshold value of $r = 0.5$.

5 (0.50 points) Repeat the previous (item) exercise with the "least strict" stopping criterion, that is, by partitioning a node whenever it does not result in empty children.

6 (0.25 points) Suppose that the different objects in our problem are equally probable, that is, $P(\mathbf{x}_1) = P(\mathbf{x}_2) = P(\mathbf{x}_3) = P(\mathbf{x}_4) = 0.25$. From the classifiers obtained above, is there any classifier whose (theoretical) classification error is better than the others? Provide a brief explanation.

1 $\mathbf{w}_0 = \mathbf{w}_1 = (0\ 0\ 0)^t$, $\alpha = 1$ y $b = 0.1$.

$$
\begin{aligned}
&g_0(\mathbf{x}_1) = (0\ 0\ 0)(1\ 0\ 0)^t = 0\\
&g_1(\mathbf{x}_2) = (0\ 0\ 0)(1\ 0\ 0)^t = 0\\
&g_1(\mathbf{x}_1) + b > g_0(\mathbf{x}_2)?\quad \text{Sí}\\
&\to \mathbf{w}_1 = \mathbf{w}_1 - \mathbf{x}_1 = (0\ 0\ 0)^t - (1\ 0\ 0)^t = (-1\ 0\ 0)^t\\
&\to \mathbf{w}_0 = \mathbf{w}_0 + \mathbf{x}_1 = (0\ 0\ 0)^t + (1\ 0\ 0)^t = (1\ 0\ 0)^t\\
&g_1(\mathbf{x}_2) = (-1\ 0\ 0)(1\ 0\ 1)^t = -1\\
&g_0(\mathbf{x}_2) = (1\ 0\ 0)(1\ 0\ 1)^t = 1\\
&g_0(\mathbf{x}_2) + b > g_1(\mathbf{x}_2)?\quad \text{Sí}\\
&\to \mathbf{w}_0 = \mathbf{w}_0 - \mathbf{x}_2 = (1\ 0\ 0)^t - (1\ 0\ 1)^t = (0\ 0\ -1)^t\\
&\to \mathbf{w}_1 = \mathbf{w}_1 + \mathbf{x}_2 = (-1\ 0\ 0)^t + (1\ 0\ 1)^t = (0\ 0\ 1)^t\\
&g_1(\mathbf{x}_3) = (0\ 0\ 1)(1\ 1\ 0)^t = 0\\
&g_0(\mathbf{x}_3) = (0\ 0\ -1)(1\ 1\ 0)^t = 0\\
&g_0(\mathbf{x}_3) + b > g_1(\mathbf{x}_3)?\quad \text{Sí}\\
&\to \mathbf{w}_0 = \mathbf{w}_0 - \mathbf{x}_3 = (0\ 0\ -1)^t - (1\ 1\ 0)^t = (-1\ -1\ -1)^t\\
&\to \mathbf{w}_1 = \mathbf{w}_1 + \mathbf{x}_3 = (0\ 0\ 1)^t + (1\ 1\ 0)^t = (1\ 1\ 1)^t\\
&g_0(\mathbf{x}_4) = (-1\ -1\ -1)(1\ 1\ 1)^t = -3\\
&g_1(\mathbf{x}_4) = (1\ 1\ 1)(1\ 1\ 1)^t = 3\\
&g_1(\mathbf{x}_4) + b > g_0(\mathbf{x}_4)?\quad \text{Sí}\\
&\to \mathbf{w}_1 = \mathbf{w}_1 - \mathbf{x}_4 = (1\ 1\ 1)^t - (1\ 1\ 1)^t = (0\ 0\ 0)^t\\
&\to \mathbf{w}_0 = \mathbf{w}_0 + \mathbf{x}_4 = (-1\ -1\ -1)^t + (1\ 1\ 1)^t = (0\ 0\ 0)^t
\end{aligned}
$$

The final weight vectors are null vectors, same as the initial vectors.

2 No. The set of training samples **IS NOT** linearly separable. Best case is to have three samples correctly classified.

3 No, same reason as in the above item.

4 The impurity of the root node is 1. There are two possible splits for the root node. For both splits, we generate child nodes that contain one object of each class so the impurity of the child nodes is also 1. Therefore, since $\Delta\mathcal{I}(j, 0.5, root) = 0 < \epsilon$ for both splits, we declare the root node as terminal node and no partition is applied.

5 The DCT algorithm generates a complete binary tree at depth level 2 where each leaf node contains a single sample.

6 If $P(\mathbf{x}_1) = P(\mathbf{x}_2) = P(\mathbf{x}_3) = P(\mathbf{x}_4) = 0.25$, the error probability is the average probability of error of the posterior probability:

$$
\begin{aligned}
P(error) &= P(error, \mathbf{x}_1) + P(error, \mathbf{x}_2) + P(error, \mathbf{x}_2) + P(error, \mathbf{x}_2)\\
&= P(\mathbf{x}_1)\, P(error \mid \mathbf{x}_1) + P(\mathbf{x}_2)\, P(error \mid \mathbf{x}_2) + P(\mathbf{x}_3)\, P(error \mid \mathbf{x}_3) + P(\mathbf{x}_4)\, P(error \mid \mathbf{x}_4)\\
&= 0.25 \cdot (P(error \mid \mathbf{x}_1) + P(error \mid \mathbf{x}_2) P(error \mid \mathbf{x}_3) + P(\mathbf{x}_4)\, P(error \mid \mathbf{x}_4))
\end{aligned}
$$

Only the learning tree obtained in the above question classifies the four samples correctly, so the probability of error of this tree is zero. As for the other classifiers, they would wrongly classify one or more of the four samples.