Intelligent Systems: Exam Block 2

ETSINF, Universitat Politècnica de València, January 13, 2016

Surname(s): [ ] **Name:** [ ]

Group: ☐ **3A** ☐ **3B** ☐ **3C** ☐ **3D** ☐ **3E** ☐ **3F** ☐ **RE1** ☐ **RE2**

Tick only one choice among the given options.

1 $\boxed{\text{D}}$ Which of the following expressions is **CORRECT**?

A) $P(x \mid y) = \frac{1}{P(z)} \sum_x P(x,y,z)$

B) $P(x \mid y) = \frac{1}{P(z)} \sum_z P(x,y,z)$

C) $P(x \mid y) = \frac{1}{P(y)} \sum_x P(x,y,z)$

D) $P(x \mid y) = \frac{1}{P(y)} \sum_z P(x,y,z)$

2 $\boxed{\text{A}}$ A physician knows that:

- The meningitis disease causes neck stiffness in the 70 % of the cases.
- The prior probability that a patient suffers from meningitis is 1 / 100 000.
- The prior probability that a patient has neck stiffness is 1 %.

Based on the above knowledge, the probability P that a patient who has neck stiffness suffers from meningitis is:

A) $0.000 \leq P < 0.001$ $\quad P = P(m \mid r) = \frac{P(m)\,P(r|m)}{P(r)} = \frac{1/100\,000 \cdot 70/100}{1/100} = 0.0007$

B) $0.001 \leq P < 0.002$

C) $0.002 \leq P < 0.003$

D) $0.003 \leq P$

3 $\boxed{\text{D}}$ Let's consider a typical classification problem in C classes and objects represented through D-dimensional real feature vectors. In general, we can say that it it more difficult to find an accurate classifier when ...

A) the values of C and D are smaller
B) the value of C is smaller and the value of D is larger
C) the value of C is larger and the value of D is smaller
D) the values of C and D are larger

4 $\boxed{\text{B}}$ We have learnt two different classifiers, c_A and c_B, for a classification problem. The probability of error of c_A has been empirically estimated from 100 test samples, obtaining an empirical estimate of error $\hat{p}_A = 0.10$ (10 %). Similarly, the probability of error of c_B has been empirically estimated but with a set of 200 test samples, obtaining an empirical estimate of error of 10 %, too ($\hat{p}_B = 0.10$). Based on these estimations, we can affirm with a 95 % of confidence that:

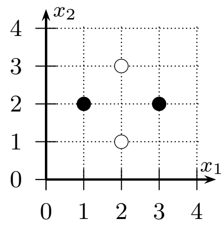
A) The confidence intervals of $\hat{p}_A$ y $\hat{p}_B$ are identical.

B) The confidence interval of $\hat{p}_A$ is larger than the one of $\hat{p}_B$. $\quad I_A = \hat{p}_A \pm 1.96 \sqrt{\frac{\hat{p}_A\,(1-\hat{p}_A)}{100}} = 0.10 \pm 0.06$

C) The confidence interval of $\hat{p}_B$ is larger than the one of $\hat{p}_A$. $I_B = \hat{p}_B \pm 1.96\sqrt{\frac{\hat{p}_B\,(1-\hat{p}_B)}{200}} = 0.10 \pm 0.04$

D) In this case, the confidence intervals of $\hat{p}_A$ and $\hat{p}_B$ are irrelevant because the estimate of error is the same.

5 $\boxed{\text{C}}$ The figure on the right shows 4 two-dimensional samples classified in 2 classes: $\circ$ and $\bullet$. If we apply the Perceptron algorithm with initial weight vectors $\mathbf{a}_\circ = (0,1,0)^t$ and $\mathbf{a}_\bullet = (0,0,1)^t$, a learning rate $\alpha > 0$ and a margin b, indicate which assertion is **CORRECT**:



A) The algorithm will converge for some $b > 0$
B) The algorithm only converges if $b \leq 0$
C) If $b > 0$ there is no convergence but, by adjusting α, we can obtain good solutions after a finite number of iterations with respect to the probability of classification error (with 25 % of misclassification error)
D) The algorithm is not applicable in this case because the classes are non-linearly separable.

6 $\boxed{\text{B}}$ Which is the number of errors of a minimum-error linear classifier for the training samples of the above question?

A) 0
B) 1
C) 2
D) 3

7 $\boxed{\text{B}}$ Given a linear classifier of two classes $\circ$ and $\bullet$ with weight vectors $\mathbf{a}_\circ = (3,1,1)^t$ and $\mathbf{a}_\bullet = (1,2,1)^t$, respectively (the first component is the threshold or independent term of the linear function), which assertion is **CORRECT**?

A) There are four decision regions because there are two weight vectors and it is a two-dimensional representation space

B) The weight vectors $\mathbf{a}_\circ = (2,-2,-2)^t$ and $\mathbf{a}_\bullet = (-2,0,-2)^t$ define the same decision boundary than the weight vectors given in the question statement The decision boundary equation is: $\mathbf{a}_\circ^t \mathbf{y} = \mathbf{a}_\bullet^t \mathbf{y}$. In both cases, we have: $y_1 = 2$.

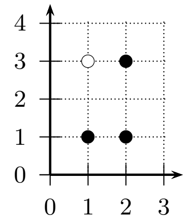
C) The weight vectors $\mathbf{a}_\circ = (1,2,1)^t$ y $\mathbf{a}_\bullet = (3,1,1)^t$ define an equivalent classifier to the one given in the statement Opposed decision regions.

D) The decision boundary is defined as a plane in $\mathbb{R}^3$ because the weight vectors are three-dimensional

8 $\boxed{\text{D}}$ Let's assume we apply the Decision Classification Tree (DCT) algorithm for a two-class problem, A and B. The DCT algorithm reaches a node t which includes two data: one sample that belongs to class A and the other belongs to class B. The entropy impurity of t, $\mathcal{I}(t)$, is:

A) $\mathcal{I}(t) < 0.0$
B) $0.0 \leq \mathcal{I}(t) < 0.5$
C) $0.5 \leq \mathcal{I}(t) < 1.0$
D) $1.0 \leq \mathcal{I}(t)$ $\mathcal{I}(t) = -\hat{P}(A \mid t)\,\log_2 \hat{P}(A \mid t) - \hat{P}(B \mid t)\,\log_2 \hat{P}(B \mid t) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$

9 $\boxed{D}$ The figure on the right shows a two-cluster partition of four two-dimensional data (represented by the symbols $\bullet$ and $\circ$). The Sum of Square Errors (SSE) of this partition is $J = \frac{30}{9}$. The transfer of the point $(2,3)^t$ from cluster $\bullet$ to $\circ$ leads to an increase in the SSE, ΔJ, such that:



A) $\Delta J > 0$

B) $0 \geq \Delta J > -1$

C) $-1 \geq \Delta J > -2$

D) $-2 \geq \Delta J$ $\qquad \Delta J = -\frac{21}{9} = -2.33 \quad (J = \frac{30}{9} \to J = 1)$

10 $\boxed{B}$ Two well-known versions of the K-means algorithm are the *Duda and Hart* (DH) version and the "popular" version. Assuming both versions are applied in the same initial partition, indicate which of the following assertions is TRUE:

A) Both versions will get the same optimized partition

B) The DH version will get a partition which cannot be further improved with the "popular" version

C) The "popular" version will get a partition which cannot be further improved with the DH version

D) The final partition obtained with DH would could be further improved with the "popular" version and viceversa

11 $\boxed{A}$ Given the Markov model M_A of the question 12, the approximated probability of the string "`bba`" calculated with Viterbi is:

A) 0.003200 $\tilde{P}(bba, q_1q_2q_3 = 111 \mid M_A) = 0.2 \cdot 0.8 \cdot 0.5 \cdot 0.8 \cdot 0.5 \cdot 0.2 \cdot 0.5 = 0.0032$

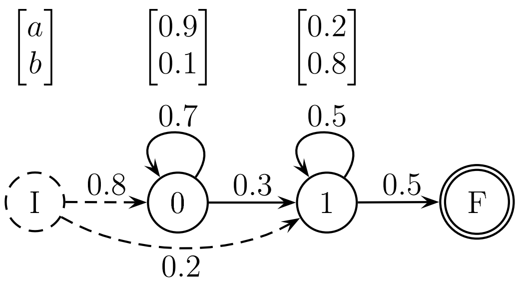
B) 0.004328

C) 0.006400

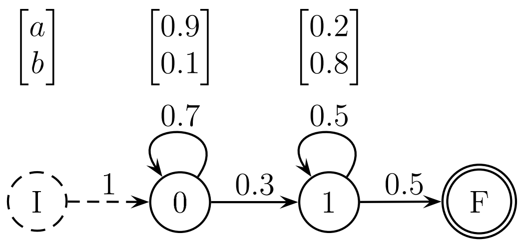
D) None of the above options are correct.

12 $\boxed{B}$ We have two equiprobable classes, A and B, for classifying strings of symbols in the alphabet $\Sigma = \{a, b\}$. The conditional probabilities of the classes are characterized by the Markov models

Model M_A: $P(x \mid A) = P(x \mid M_A)$



Model M_B: $P(x \mid B) = P(x \mid M_B)$



Indicate the **CORRECT** option if we want to classify the string "`bba`" by minimum classification error:

A) Either A or B because both classes are equiprobable

B) Class A. $\hat{c} = \arg\max_c P(c \mid \text{"}bba\text{"}) = \arg\max_c P(c)P(\text{"}bba\text{"} \mid c) = \arg\max_c P(\text{"}bba\text{"} \mid c)$

C) Class B. $P(\text{"}bba\text{"} \mid A) \approx \tilde{P}(\text{"}bba\text{"}|A) = 0.0032 \gg P(\text{"}bba\text{"} \mid B) \approx \tilde{P}(\text{"}bba\text{"}|B) = 0.0012 \to \hat{c} = A$

D) It cannot be determined because M_B does not satisfy the normalization conditions.

13 $\boxed{C}$ Given the Markov model M_A of the question 12, if we apply the *Forward* algorithm to the string "`bba`", mark the **CORRECT** expression:

A) $\alpha(q=1,t=3) = \alpha(q=0,t=2) \cdot A_{01} \cdot B_{1a}$

B) $\alpha(q=1,t=3) = \alpha(q=1,t=2) \cdot A_{11} \cdot B_{1a}$

C) $\alpha(q=1,t=3) = \alpha(q=0,t=2) \cdot A_{01} \cdot B_{1a} + \alpha(q=1,t=2) \cdot A_{11} \cdot B_{1a}$

D) $\alpha(q=1,t=3) = \alpha(q=0,t=2) \cdot A_{01} \cdot B_{1a} \cdot \alpha(q=1,t=2) \cdot A_{11} \cdot B_{1a}$

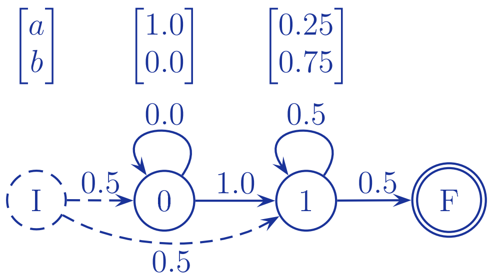
14 $\boxed{D}$ Given the Markov model M_A of the question 12, if we apply ONE iteration of Viterbi re-estimation algorithm with the strings "`bba`" and "`ab`", which assertion is **CORRECT**?:

A) $\pi_0 = 1$

B) No changes are produced in the model.

C) All the transition probabilities change their value.

D) Some of the transition and emission probabilities of state 0 are null.



15 $\boxed{B}$ Given a Markov model with states $Q = \{0, 1, F\}$ and alphabet $\Sigma = \{a, b\}$ initialized through a linear segmentation with the strings "`bbaa`" y "`ab`", indicate the **CORRECT** choice:

A) Some emission probabilities are null.

B) It holds that $A_{00} = A_{11}$ and $A_{01} = A_{1F}$

C) It holds that $\pi_0 = \pi_1$

D) It holds that $B_{0a} = B_{1a}$