Intelligent Systems: Exam Block 2

ETSINF, Universitat Politècnica de València, January 18, 2017

Surname(s): | | **Name:** | |

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

Tick only one choice among the given options.

1 $\boxed{C}$ Which of the following expressions is **INCORRECT**?

A) $P(x \mid y) = \dfrac{P(x,y)}{\sum_z P(y \mid z)\ P(z)}$

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

C) $P(x \mid y) = \dfrac{\sum_z P(x,z)}{P(y)}$

D) $P(x \mid y) = \dfrac{P(y \mid x)\ P(x)}{P(y)}$

2 $\boxed{B}$ We have two bags of apples. The first one has 3 red apples and 5 green apples. The second bag contains 2 red apples, 2 green apples and 1 yellow apple. We randomly pick one bag and, subsequently, a randomly apple from such a bag. Let's suppose that both bags are equally probable of being chosen and that, given one particular bag, its apples are also equally probable of being chosen. Assuming we pick a red apple, which is the probability P that this apple belongs to the first bag?

A) $0.00 \leq P < 0.25$

B) $0.25 \leq P < 0.50$

C) $0.50 \leq P < 0.75$

D) $0.75 \leq P$

$$\begin{aligned} P = P(B=1 \mid C=r) &= \frac{P(B{=}1)P(C{=}r|B{=}1)}{P(C{=}r)} \\ &= \frac{P(B{=}1)P(C{=}r|B{=}1)}{P(B{=}1)P(C{=}r|B{=}1)+P(B{=}2)P(C{=}r|B{=}2)} \\ &= \frac{\frac{1}{2}\cdot\frac{3}{8}}{\frac{1}{2}\cdot\frac{3}{8}+\frac{1}{2}\cdot\frac{2}{5}} = \frac{15}{31} = 0.4839 \end{aligned}$$

3 $\boxed{A}$ Let x be an object (feature vector or string) we wish to classify among C possible classes. Indicate which of the following expressions **IS NOT** a minimum-error classifier.

A) $c(x) = \underset{c=1,\dots,C}{\arg\max}\ p(x \mid c)$

B) $c(x) = \underset{c=1,\dots,C}{\arg\max}\ p(x, c)$

C) $c(x) = \underset{c=1,\dots,C}{\arg\max}\ \log p(x, c)$

D) $c(x) = \underset{c=1,\dots,C}{\arg\max}\ P(c \mid x)$

4 $\boxed{A}$ Let be a classification problem of two classes in $\mathbb{R}^2$. We have a classifier made up of two discriminant linear functions with weight vectors $\mathbf{a}_\circ = (-1,1,2)^t$ and $\mathbf{a}_\bullet = (1,1,1)^t$. Indicate the decision regions defined by this classifier.

A) $R_\circ = \{\mathbf{x} \in \mathbb{R}^2 : x_2 > 2\}$ y $R_\bullet = \{\mathbf{x} \in \mathbb{R}^2 : x_2 < 2\}$

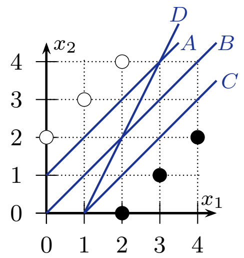
B) $R_\circ = \{\mathbf{x} \in \mathbb{R}^2 : x_1 > 2\}$ y $R_\bullet = \{\mathbf{x} \in \mathbb{R}^2 : x_1 < 2\}$

C) $R_\circ = \{\mathbf{x} \in \mathbb{R}^2 : x_1 < 2\}$ y $R_\bullet = \{\mathbf{x} \in \mathbb{R}^2 : x_1 > 2\}$

D) $R_\circ = \{\mathbf{x} \in \mathbb{R}^2 : x_2 < 2\}$ y $R_\bullet = \{\mathbf{x} \in \mathbb{R}^2 : x_2 > 2\}$

$g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \rightarrow x_2 = 2 \ \wedge\ g_\circ((0,0)^t) < g_\bullet((0,0)^t)$

5 $\boxed{\text{B}}$ The figure on the right shows 6 two-dimensional samples of two classes ($\circ$ and $\bullet$). After applying the Perceptron algorithm with different values of the parameter b, we obtain the 4 classifiers that appear in the response choices. Which classifier returns the most centered decision boundary and consequently the lowest expected error?



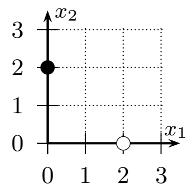
A) $\mathbf{a}_\circ = (-1, 1, 2)^t$ y $\mathbf{a}_\bullet = (\ \ 0, 2, 1)^t$ $\quad g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \to x_2 = x_1 + 1$

B) $\mathbf{a}_\circ = (\ \ 1, 1, 2)^t$ y $\mathbf{a}_\bullet = (\ \ 1, 2, 1)^t$ $\quad g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \to x_2 = x_1$

C) $\mathbf{a}_\circ = (\ \ 1, 1, 2)^t$ y $\mathbf{a}_\bullet = (\ \ 0, 2, 1)^t$ $\quad g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \to x_2 = x_1 - 1$

D) $\mathbf{a}_\circ = (\ \ 1, 1, 1)^t$ y $\mathbf{a}_\bullet = (-1, 3, 0)^t$ $\quad g_\circ(\mathbf{x}) = g_\bullet(\mathbf{x}) \to x_2 = 2x_1 - 2$

6 $\boxed{\text{C}}$ The figure on the right shows 2 bidimensional samples of two classes ($\mathbf{x}_1$,$\circ$) and ($\mathbf{x}_2$,$\bullet$). Given the weights $\mathbf{a}_\circ = (0, 1, 0)^t$ and $\mathbf{a}_\bullet = (1, 0, 0)^t$, if we apply one iteration of the Perceptron algorithm using only the sample $\mathbf{x}_1$, which is the minimum value of the margin b for which the weight vectors are updated?



A) $b = 0.5$

B) $b = 1.0$

C) $b = 1.5$ $\quad g_\circ(\mathbf{x}_1) = 2 \quad g_\bullet(\mathbf{x}_1) = 1 \quad$ if $(g_\bullet(\mathbf{x}_1) + b > g_\circ(\mathbf{x}_1))$

D) None of the above

7 $\boxed{\text{D}}$ We have learnt a classifier for a classification problem and we have empirically estimated the error for a given set of training samples, obtaining the 95 % confidence interval of the probability of error. Which option would allow us to reduce the size of this interval?

A) Decrease significantly the size of the training set

B) Keep the training set and re-train the classifier with the *correct* version of the K-means algorithm (Duda & Hart algorithm)

C) Keep the training set and re-train the classifier with the *popular* version of the K-means algorithm

D) Increase significantly the size of the training set

8 $\boxed{\text{A}}$ We have 6 three-dimensional samples for a classification problem of three classes (A, B and C) (see table). If we apply the decision tree learning algorithm on these data, which is the number (N) of different splits to be explored in the root node? (Note: discard the splits that yield empty nodes)

n	1	2	3	4	5	6
x_{n1}	0	1	0	1	0	1
x_{n2}	1	1	2	2	3	3
x_{n3}	0	2	0	3	2	3
c_n	A	A	B	B	C	C

A) $0 \le N \le \ 5$ $\quad \{(1,0), (2,1), (2,2), (3,0), (3,2)\}$

B) $\ 5 < N \le 10$

C) $10 < N \le 20$

D) An infinitely number of splits.

9 $\boxed{\text{D}}$ Let's suppose we apply the decision tree learning algorithm for a classification problem of four classes $\mathcal{C} = \{1, 2, 3, 4\}$. The algorithm reaches a node t that contains 8 data: four data of class 1, two data of class 2, one data of class 3 and one data of class 4. The impurity of t, $\mathcal{I}(t)$, measured as the entropy impurity of the posterior probability of the four classes in t is:

A) $0.00 \le \mathcal{I}(t) < 0.25$

B) $0.25 \le \mathcal{I}(t) < 0.50$

C) $0.50 \le \mathcal{I}(t) < 0.75$

D) $0.75 \leq \mathcal{I}(t)$ $\quad \mathcal{I}(t) = -\sum_{c=1}^{4} \hat{P}(c \mid t) \log_2 \hat{P}(c \mid t) = -\frac{4}{8}\log_2\frac{4}{8} - \frac{2}{8}\log_2\frac{2}{8} - 2\,\frac{1}{8}\log_2\frac{1}{8} = \frac{7}{4} = 1.75$

10 $\boxed{B}$ Indicate which of the following assertions about Supervised Learning (SL) and Unsupervised Learning (UL) is **CORRECT**:

A) Both SL and UL require class unlabeled training data

B) UL requires class unlabeled data and SL requires class labeled data

C) UL requires class labeled data and SL requires class unlabeled data

D) Both SL and UL require class labeled training data

11 $\boxed{D}$ Consider the *correct* version of Duda & Hart (DH) and the *popular* version (PV) of the K-means algorithm. Both optimize the sum of squared errors (SSE) but their result differ because:

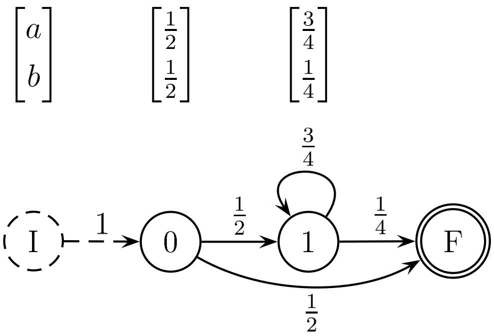
A) DH minimizes SSE and PV maximizes SSE

B) DH maximizes SSE and PV minimizes SSE

C) Both maximize SSE but DH yields better solutions than PV

D) None of the above

12 $\boxed{C}$ In the Hidden Markov Model of the right figure, $P_M(a) = P_M(b) = \frac{1}{4}$. Which is the value of $S = \sum_x P_M(x)$, where x is any string formed by two or more symbols?



A) $0 \leq S < \frac{1}{4}$.

B) $\frac{1}{4} \leq S < \frac{2}{4}$.

C) $\frac{2}{4} \leq S < \frac{3}{4}$.

D) $\frac{3}{4} \leq S \leq 1$.

13 $\boxed{B}$ Let M be a Hidden Markov Model and x a string such that $P_M(x) > 0$. It **ALWAYS HOLDS** that:

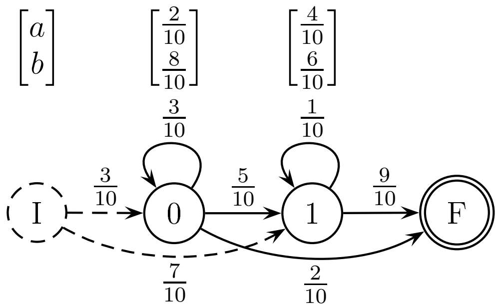
A) There is a single sequence of states that generates x with maximum probability

B) There is a single Viterbi approximation to $P_M(x)$

C) There are several sequences of states that generate x with maximum probability

D) There are several Viterbi approximations to $P_M(x)$

14 $\boxed{D}$ We have a two-class (A and B) classification problem of objects that represent strings in the alphabet $\Sigma = \{a, b\}$. The conditional probability functions of the two classes are given by the Hidden Markov Models M_A and M_B. We know that $P(A) = 0.45$, $P(ba|A) = P_{M_A}(ba) = 0.0612$ and that M_B is the MM shown in the figure on the right, where $P(ba|B) = P_{M_B}(ba)$. Assuming we want to classify the string "`ba`" by minimum classification error, mark the **CORRECT** assertion:



$\hat{c} = \arg\max_c P(c)P(ba \mid c)$

$P(A)P(ba \mid A) = 0.45 \cdot 0.0612$

$P(B)P(ba \mid B) = 0.55 \cdot 0.0612$

$\hat{c} = B$

A) Given the provided data, it is not possible to determine the class of '`ba`".

B) The string '`ba`" belongs indistinctly to class A or B because $P_{M_A}(ba) = P_{M_B}(ba)$.

C) The string '`ba`" belongs to class A.

D) The string '`ba`" belongs to class B.

15 $\boxed{A}$ Given the Markov Model M_B of the above question, after applying one iteration of Vitebi re-estimation with the strings "`ba`", "`b`" and "`aa`", show the **CORRECT** result:

A) $A_{01} = A_{1F} = 1$

B) $B_{0a} = B_{1a} = \frac{1}{2}$

C) $\pi_0 = \frac{1}{3}$

D) $\pi_1 = \frac{2}{3}$

A	0	1	F
0	0	1	0
1	0	0	1

B	a	b
0	$\frac{1}{2}$	$\frac{1}{2}$
1	$\frac{2}{3}$	$\frac{1}{3}$

$\pi_0 = \frac{2}{3}$ $\pi_1 = \frac{1}{3}$