Intelligent Systems
Exercises Block 2, Chapter 1
Probabilistic Reasoning

Escuela Técnica Superior de Informática
Dep. de Sistemas Informáticos y Computación
Universitat Politècnica de València

6 de noviembre de 2022

Questions

1 $\boxed{\text{B}}$ (Exam January 15, 2014) Given the joint probability of two random variables X and Y, the conditional probability $P(Y = y \mid X = x)$ can be calculated as:

A) $P(y \mid x) \;=\; 1 \,/\, P(x, y)$
B) $P(y \mid x) \;=\; P(x, y) \,/\, \sum_{y'} P(x, y')$
C) $P(y \mid x) \;=\; \sum_{x'} P(x', y) \,/\, \sum_{y'} P(x, y')$
D) $P(y \mid x) \;=\; \sum_{x'} P(x', y) \cdot \sum_{y'} P(x, y')$

2 $\boxed{\text{A}}$ (Exam January 15, 2014) In a binary decision problem ($D = \{0, 1\}$), let y be a fact or data and $d^\star(y) = 0$ be the decision of optimal classification (minimum error classification). Indicate which of the following expressions is ***not correct*** to determine the minimum probability of error for y:

A) $P_\star(\text{error} \mid Y = y) \;=\; 1 - P(D = 1 \mid Y = y)$
B) $P_\star(\text{error} \mid Y = y) \;=\; 1 - P(D = 0 \mid Y = y)$
C) $P_\star(\text{error} \mid Y = y) \;=\; P(D = 1 \mid Y = y)$
D) $P_\star(\text{error} \mid Y = y) \;=\; 1 - \text{máx}_d\, P(D = d \mid Y = y)$

3 $\boxed{\text{D}}$ (Exam January 15, 2014) In a differential diagnosis between *Flu* and *Cold*, we know that the relative occurrence of *Flu* with respect to *Cold* is 30 %. We know the following distribution of fever values in Celsius degrees:

$t(^oC)$	36	37	38	39	40
$P(T = t \mid D = \text{FLU})$	0.05	0.10	0.20	0.30	0.35
$P(T = t \mid D = \text{COLD})$	0.10	0.30	0.40	0.15	0.05

The conditional (posterior) probability that a patient has *Flu* given he has a fever of 38^oC is:

A) greater than 0.8
B) lower than 0.1
C) between 0.3 and 0.6
D) lower than the probability that the patient has *Cold* with the same fever of 38^oC

4 $\boxed{\text{D}}$ (Exam January 28, 2014) In a classification experiment with 300 test samples, 15 wrong decisions were found. With a 95 % of confidence, we can affirm that the true probability of error is:

A) $P(\text{error}) = 5\,\% \pm 0.3\,\%$
B) $P(\text{error}) = 0.05 \pm 0.3$
C) $P(\text{error}) = 0.05$, exactly
D) $P(\text{error}) = 0.05 \pm 0.03$

$$0.05 \pm 1.96\sqrt{\frac{0.05\,0.95}{300}} = 0.05 \pm 0.03 \quad (5\,\% \pm 3\,\%)$$

5 $\boxed{\text{B}}$ (Exam January 28, 2014) In a differential diagnosis between *Flu* and *Cold*, we know that the relative occurrence of *Flu* with respect to *Cold* is 30 %. We know the following distribution of fever values in Celsius degrees:

$t(^oC)$	36	37	38	39	40
$P(T = t \mid D = \text{FLU})$	0.05	0.10	0.20	0.30	0.35
$P(T = t \mid D = \text{COLD})$	0.10	0.30	0.40	0.15	0.05

$$P(\text{FLU} \mid 37) = \frac{\frac{30}{130}\,0.10}{\frac{30}{130}\,0.10 + \frac{100}{130}\,0.30} = \frac{1}{11}$$

The most probable diagnosis for a patient that has a fever of 37^oC is:

A) *Flu*
B) *Cold*
C) There is a tie between the two diagnosis.
D) The given probabilities are not correct because they don't sum up 1; therefore, a diagnosis cannot be made.

6 $\boxed{\text{C}}$ (January 13, 2015) Regarding the Bayes' rule, which of the following expressions is ***not correct***?

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

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

7 $\boxed{\text{B}}$ (January 13, 2015) The commercial assessment of the 300 movies screened in a cinema over the last year was *success* for 120 movies and *failure* for the rest of the movies. We know the distribution of the movie genres given its commercial assessment:

g	ROMANCE	COMEDY	INTRIGUE
$P(G = g \mid A = \text{SUCCESS})$	0.30	0.35	0.35
$P(G = g \mid A = \text{FAILURE})$	0.20	0.50	0.30

Which is the most probable commercial assessment for an intrigue film?

A) *Success*
B) *Failure* $\quad P(V = \text{FRACASO} \mid \text{G=INTRIGA}) = 0.5625$
C) Both commercial assessments have the same probability
D) It is impossible to determine the prediction with the available data

8 $\boxed{\text{D}}$ (January 13, 2015) In a classification problem in three classes, $C = \{a, b, c\}$, let y be a fact or data. The decision of optimal classification for y is class a with a posterior probability of 0.40. Which of the following assertions is ***incorrect***?

A) $P(C = a \mid Y = y) \;\le\; P(C = b \mid Y = y) + P(C = c \mid Y = y)$
B) $P_\star(\text{error} \mid Y = y) \;=\; P(C = b \mid Y = y) + P(C = c \mid Y = y)$
C) $P_\star(\text{error} \mid Y = y) \;=\; 1 - P(C = a \mid Y = y)$
D) $P_\star(\text{error} \mid Y = y) \;=\; 1 - \text{máx}_{d \in \{b,c\}}\, P(C = d \mid Y = y)$

9 $\boxed{\text{D}}$ (January 26, 2015) Let X, Y and Z be three random variables. X and Y are *conditionally independent* given Z if and only if

$$P(X = x, Y = y \mid Z = z) \;=\; P(X = x \mid Z = z)\, P(Y = y \mid Z = z) \qquad \text{for all } x,\ y \text{ y } z.$$

If the above condition holds, $P(Z = z \mid X = x, Y = y)$ can be calculated as ...:

A) $P(Z = z \mid X = x, Y = y) \;=\; \dfrac{P(X = x, Y = y, Z = z)}{P(X = x, Y = y)}$

B) $P(Z = z \mid X = x, Y = y) \;=\; \dfrac{P(Z = z)\, P(X = x, Y = y \mid Z = z)}{P(X = x, Y = y)}$

C) $P(Z = z \mid X = x, Y = y) \;=\; \dfrac{P(Z = z)\, P(X = x \mid Z = z)\, P(Y = y \mid Z = z)}{P(X = x, Y = y)}$

D) The three above answers are all correct to calculate $P(Z = z \mid X = x, Y = y)$.

10 $\boxed{C}$ (January 26, 2015) Let be a classification problem in three classes, $C = \{a, b, c\}$, where the number of samples of class a is 100, the number of samples of class b is 100, and the number of samples of class c is 100, and let y be a fact or data. The decision of optimal classification for y is class a with a posterior probability of 0.50. Which of the following assertions is ***correct***?

A) $P(C = a \mid Y = y) \ > \ P(C = b \mid Y = y) + P(C = c \mid Y = y)$

B) $P(Y = y \mid C = a) \ = \ \dfrac{0.5\ P(C = a)}{P(Y = y)}$

C) $P(Y = y \mid C = a) \ = \ P(Y = y \mid C = b) + P(Y = y \mid C = c)$

D) None of the above.

11 $\boxed{D}$ (January, 2016) Which of the following expressions is **CORRECT**?

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

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

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

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

12 $\boxed{A}$ (January, 2016) 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) $0.003 \leq P$

13 $\boxed{D}$ (January 2016) Which of the following assertions is **TRUE**?

A) $P(x, y) \ = \ \sum_z P(x)\,P(y)\,P(z).$

B) $P(x, y) \ = \ \sum_z P(x)\,P(y \mid z).$

C) $P(x, y) \ = \ \sum_z P(x \mid z)\,P(y \mid z)\,P(z).$

D) $P(x, y) \ = \ \sum_z P(x, y \mid z)\,P(z).$ $\quad P(x, y) = \sum_z P(x, y, z) = \sum_z P(x, y \mid z)\,P(z)$

14 $\boxed{A}$ (January 2016) An entomologist discovers a rare subspecies of beetle, due to the pattern of his back. In this rare subspecies, 98 % of the specimen have this pattern. In the common subspecies, 5 % of the specimen have this pattern. The rare subspecies represents 0.1 % of the population. The probability P that a beetle with the pattern of his back belongs to the rare subspecies is:

A) $0.00 \leq P < 0.05.$ $\quad P = P(r \mid p) = \frac{P(r)\,P(p|r)}{P(p)} = \frac{P(r)\,P(p|r)}{P(r)\,P(p|r)+P(c)\,P(p|c)} = \frac{1/1000 \cdot 98/100}{1/1000 \cdot 98/100 + 999/1000 \cdot 5/100} = \frac{98}{5093} = 0.0192$

B) $0.05 \leq P < 0.10.$

C) $0.10 \leq P < 0.20.$

D) $0.20 \leq P.$

15 $\boxed{\text{C}}$ (January 2017) Let x be an object (represented with a feature vector or string of symbols) that we want to classify in one among C possible classes. Indicate which of the following expressions **DOES NOT** classify x by minimum classification error:

A) $c(x) = \arg\max_{c=1,\dots,C} \log_2 p(c \mid x)$
B) $c(x) = \arg\max_{c=1,\dots,C} \log_{10} p(c \mid x)$
C) $c(x) = \arg\max_{c=1,\dots,C} a\,p(c \mid x) + b$ being a and b two real constants
D) $c(x) = \arg\max_{c=1,\dots,C} p(c \mid x)^3$

16 $\boxed{\text{C}}$ (January 2016) 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) $P(x \mid y) = \dfrac{P(y \mid x)\, P(x)}{P(y)}$

17 $\boxed{\text{B}}$ (January 2017) 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) $0.75 \leq P$

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

18 $\boxed{\text{A}}$ (January 2017) 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) $c(x) = \arg\max_{c=1,\dots,C} P(x \mid c)$

B) $c(x) = \arg\max_{c=1,\dots,C} P(x, c)$

C) $c(x) = \arg\max_{c=1,\dots,C} \log P(x, c)$

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

19 $\boxed{\text{C}}$ (January 2017) 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 $P(X)$.

20 $\boxed{\text{A}}$ (January 2018) Mark the **INCORRECT** expression.
A) $\sum_y P(x \mid y) = 1, \ \forall x$
B) $\sum_x P(x \mid y) = 1, \ \forall y$
C) $\sum_x \sum_y P(x, y) = 1$
D) $\sum_x P(x \mid u) = \sum_y P(y \mid w), \ \forall u, w$

21 $\boxed{\text{B}}$ (January 2018) We have two orange stores (1 and 2). Store 1 contains 65 % of the oranges; the rest are in store 2. We know that store 1 contains 1 % of the unsuitable oranges for human consumption, and store 2 contains 3 % of the unsuitable oranges for human consumption. Let's assume that one unsuitable orange is delivered. Which is the probability P that such an orange comes from store 1?

A) $0.00 \leq P < 0.25$

B) $0.25 \leq P < 0.50$ $\quad P = P(A=1|C=0) = \frac{P(A=1)P(C=0|A=1)}{P(C=0)} = \frac{P(A=1)P(C=0|A=1)}{P(A=1)P(C=0|A=1)+P(A=2)P(C=0|A=2)} = 0.38$

C) $0.50 \leq P < 0.75$

D) $0.75 \leq P$

22 $\boxed{\text{D}}$ (January 2018) Let $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ be an object defined by means of a sequence of N feature vectors. We want to classify $\mathbf{x}$ among one of C classes. Show which of the following expressions represents a classifier that minimizes the probability of error. ($\mathbf{x}_2^N$ denotes $\mathbf{x}_2, \ldots, \mathbf{x}_N$):

A) $c(\mathbf{x}) = \underset{c=1,\ldots,C}{\arg\max}\ p(\mathbf{x}_1 \mid c)\, p(\mathbf{x}_2^N \mid \mathbf{x}_1)$

B) $c(\mathbf{x}) = \underset{c=1,\ldots,C}{\arg\max}\ p(\mathbf{x}_1, c)\, p(\mathbf{x}_2^N \mid \mathbf{x}_1)$

C) $c(\mathbf{x}) = \underset{c=1,\ldots,C}{\arg\max}\ p(\mathbf{x}_1 \mid c)\, p(\mathbf{x}_2^N \mid \mathbf{x}_1, c)$

D) $c(\mathbf{x}) = \underset{c=1,\ldots,C}{\arg\max}\ p(\mathbf{x}_1, c)\, p(\mathbf{x}_2^N \mid \mathbf{x}_1, c)$

23 $\boxed{\text{B}}$ (January 2018) We have a 3-class classifier for $\mathbf{x} = (x_1, x_2)^t \in [0,1]^2$ with the probability distributions shown on the right table. Which is the probability of error (p_e) of the classifier?

A) $p_e < 0.35$

B) $0.35 \leq p_e < 0.45$ $\quad .1 \cdot 0 + .2 \cdot .02 + .3 \cdot .5 + .4 \cdot 2/3 = .42$

C) $0.45 \leq p_e < 0.65$

D) $0.65 \leq p_e$

x_1	x_2	$p(c=1\|\mathbf{x})$	$p(c=2\|\mathbf{x})$	$p(c=3\|\mathbf{x})$	$p(\mathbf{x})$
0	0	1.0	0.0	0.0	0.1
0	1	0.01	0.01	0.98	0.2
1	0	0.25	0.5	0.25	0.3
1	1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0.4

24 $\boxed{\text{D}}$ (January 2018) Let be a classification problem among 4 equiprobable classes, $c = 1, 2, 3, 4$. For a given object, x, we know that the Bayes classifier assigns class 1 to x and that the posterior probability of belonging to that class, $p(c = 1 \mid x)$, is 1/3. Given this information, show the **CORRECT** statement:

A) The probability of error of classifying x is lower than 1/3.

B) $p(c = 1 \mid x) \ > \ p(c = 2 \mid x) + p(c = 3 \mid x) + p(c = 4 \mid x)$.

C) $p(x) > p(x \mid c = 1)$.

D) None of the above.

25 $\boxed{\text{D}}$ (December 2018) Which of the following probability distributions **CANNOT** be deduced from the joint probability $P(x, y, z)$?

A) $P(x \mid y)$

B) $P(z \mid x, y)$

C) $P(z)$

D) Every distribution involving any combination of these variables can be deduced from $P(x, y, z)$.

26 $\boxed{\text{C}}$ (December 2018) Let be a classification problem among 4 equiprobable classes $C = \{a, b, c, d\}$ and let y be a fact or data. The decision of optimal classification for y is class a with a posterior probability of 0.30. Show the **CORRECT** statement:

A) The probability of error of classifying y is lower than 0.50.

B) $P(C = a \mid Y = y) \ > \ P(C = b \mid Y = y) + P(C = c \mid Y = y) + P(C = d \mid Y = y)$.

C) $P(Y = y \mid C = a) \ = \ \dfrac{0.3\ P(Y = y)}{0.25}$.

D) None of the above.

27 $\boxed{\text{C}}$ (December 2018) Suppose we have two boxes each containing 40 cookies. Box #1 contains 10 chocolate chip cookies and 30 plain cookies. Box #2 contains 20 cookies of each type. Let's assume we randomly pick a box, and then we randomly take a cookie from the selected box. If the picked cookie is a plain cookie, which is the probability P that the cookie is picked out of the box #1?

A) $0/4 \leq P < 1/4$.
B) $1/4 \leq P < 2/4$.
C) $2/4 \leq P < 3/4$.
D) $3/4 \leq P \leq 4/4$.

$$P(C=1 \mid G=c) = \frac{P(C=1)\,P(G=c \mid C=1)}{P(C=1)\,P(G=c \mid C=1) + P(C=2)\,P(G=c \mid C=2)} = \frac{1/2 \cdot 3/4}{1/2 \cdot 3/4 + 1/2 \cdot 1/2} = \frac{3}{5} = 0.6$$

28 $\boxed{\text{A}}$ (December 2018) Let $\mathbf{x} = (x_1, \ldots, x_D)^t$, $D > 1$, be an object represented by a D-dimensional feature vector. We want to classify $\mathbf{x}$ in one among C classes. Which of the following expressions **IS NOT** a minimum-error classifier?

A) $c(\mathbf{x}) = \underset{c=1,\ldots,C}{\arg\max}\ p(x_1 \mid c)\, p(x_2, \ldots, x_D \mid x_1, c)$
B) $c(\mathbf{x}) = \underset{c=1,\ldots,C}{\arg\max}\ p(c)\, p(x_1, \ldots, x_D \mid c)$
C) $c(\mathbf{x}) = \underset{c=1,\ldots,C}{\arg\max}\ p(c \mid x_1)\, p(x_2, \ldots, x_D \mid x_1, c)$
D) $c(\mathbf{x}) = \underset{c=1,\ldots,C}{\arg\max}\ p(x_1, c)\, p(x_2, \ldots, x_D \mid x_1, c)$

29 $\boxed{\text{B}}$ Let be a two-class ($c = 1, 2$) classification problem for objects represented in a space of 4 elements, $E = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$. The table on the right shows the (true) posterior probabilities $P(c \mid \mathbf{x})$, for every c and $\mathbf{x}$; as well as the (true) prior probability, $P(\mathbf{x})$, for every $\mathbf{x}$. Likewise, the table shows the class, $c(\mathbf{x})$, assigned to each $\mathbf{x} \in E$ by a particular classifier. On the grounds of this probabilistic knowledge, the probability of error of $c(\mathbf{x})$, ε, is:

	$P(c \mid \mathbf{x})$			
$\mathbf{x}$	$c=1$	$c=2$	$P(\mathbf{x})$	$c(\mathbf{x})$
$\mathbf{x}_1$	1	0	1/3	1
$\mathbf{x}_2$	3/4	1/4	1/4	1
$\mathbf{x}_3$	1/4	3/4	1/4	1
$\mathbf{x}_4$	1/2	1/2	1/6	2

A) $0/4 \leq \varepsilon < 1/4$.
B) $1/4 \leq \varepsilon < 2/4$. $\quad \varepsilon = \frac{1}{3} \cdot 0 + \frac{1}{4} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{3}{4} + \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{3}$
C) $2/4 \leq \varepsilon < 3/4$.
D) $3/4 \leq \varepsilon \leq 4/4$.

30 $\boxed{\text{A}}$ Let's assume we want to apply the probability of error of a Bayes classifier (Bayes probability of error), which we will denote as ε^*, to the problem of question 29. The value of ε^* is:

A) $0/4 \leq \varepsilon^* < 1/4$. $\quad \varepsilon^* = \frac{1}{3} \cdot 0 + \frac{1}{4} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{4} + \frac{1}{6} \cdot \frac{1}{2} = \frac{5}{24} = 0.2083$
B) $1/4 \leq \varepsilon^* < 2/4$.
C) $2/4 \leq \varepsilon^* < 3/4$.
D) $3/4 \leq \varepsilon^* \leq 4/4$.

31 $\boxed{\text{B}}$ Let be a two-class ($c = 1, 2$) classifier for $x \in \{0, 1\}$, where $P(c)$ and $p(x)$ are uniform probability distributions, and $p(x \mid c) = \frac{1}{c} \cdot x + (1 - \frac{1}{c}) \cdot (1 - x)$. Which is the probability of error, p_e, of a Bayes classifier?

A) $p_e < 0.25$
B) $0.25 \leq p_e < 0.50$ $\quad \frac{1}{2} \cdot \frac{1}{2} = 0.25$
C) $0.50 \leq p_e < 0.75$
D) $0.75 \leq p_e$

x	$p(x \mid c=1)$	$p(x \mid c=2)$	$p(x)$
0	0.0	$\frac{1}{2}$	$\frac{1}{2}$
1	1.0	$\frac{1}{2}$	$\frac{1}{2}$

32 $\boxed{\text{D}}$ Let $\mathbf{x}$ be an object to classify in one among C classes. Indicate the expression that **IS NOT** a minimum-error classifier (or select the last option if the first three options denote a minimum-error classifier).

A) $c(\mathbf{x}) = \underset{c=1,\ldots,C}{\arg\max}\ p(c \mid \mathbf{x})^2$.
B) $c(\mathbf{x}) = \underset{c=1,\ldots,C}{\arg\max}\ \log p(\mathbf{x}, c)$.
C) $c(\mathbf{x}) = \underset{c=1,\ldots,C}{\arg\max}\ \sqrt{p(\mathbf{x}, c)}\, /\, p(\mathbf{x})$.
D) All three above expressions denote a minimum-error classifier.

33 $\boxed{\text{A}}$ Let W, D, S be three random variables that take values from {CLE,CLO,RAI}, {DAY,NIG}, and {SAF,ACC}, respectively. The joint probability is given in the following table:

s	SAF	SAF	SAF	SAF	SAF	SAF	ACC	ACC	ACC	ACC	ACC	ACC
d	DAY	DAY	DAY	NIG	NIG	NIG	DAY	DAY	DAY	NIG	NIG	NIG
w	CLE	CLO	RAI	CLE	CLO	RAI	CLE	CLO	RAI	CLE	CLO	RAI
$P(s,d,w)$	0.30	0.20	0.07	0.13	0.10	0.06	0.01	0.01	0.03	0.02	0.02	0.05

The conditional probability $P(W = \text{RAI} | S = \text{ACC}, D = \text{DAY})$ is:

A) 0.60. $P(C = \text{RAI} \mid S = \text{ACC}, L = \text{DAY}) = P(C = \text{RAI}, S = \text{ACC}, L = \text{DAY})/P(S = \text{ACC}, L = \text{DAY})$
B) 0.03. $P(S = \text{ACC}, L = \text{DAY}) = \sum_c P(S = \text{ACC}, L = \text{DAY}, c) = 0.05$
C) 0.05. $P(C = \text{RAI} \mid S = \text{ACC}, L = \text{DAY}) = 0.03/0.05 = 0.60$
D) 0.02.

34 $\boxed{\text{C}}$ Given a classification problem among three classes for objects of type $\mathbf{x} = (x_1, x_2)^t \in \{0,1\}^2$ with the probability distribution given in the table on the right, which is the Bayes error, ε^*, of this problem?

A) $\varepsilon^* < 0.2$.
B) $0.2 \le \varepsilon^* < 0.4$.
C) $0.4 \le \varepsilon^* < 0.7$. $.2 \cdot .4 + .3 \cdot .2 + .2 \cdot .5 + .3 \cdot 2/3 = .44$
D) $0.7 \le \varepsilon^*$.

$\mathbf{x}$		$P(c \mid \mathbf{x})$			
x_1	x_2	$c{=}1$	$c{=}2$	$c{=}3$	$P(\mathbf{x})$
0	0	0.6	0.2	0.2	0.2
0	1	0.1	0.1	0.8	0.3
1	0	0.3	0.5	0.2	0.2
1	1	1/3	1/3	1/3	0.3

35 $\boxed{\text{D}}$ We have learnt a classifier for a particular problem. With a testing set of $M = 100$ samples, we have estimated:

- The estimated probability of error of the classifier: $\hat{p} = 0.10 = 10\,\%$.
- A 95 % confidence interval for this probability of error: $\hat{I} = [0.04, 0.16] = [4\,\%, 16\,\%]$.

We consider that the value of $\hat{p}$ is reasonable and that it will not significantly vary even if we used many more testing samples. However, we believe that the 95 % confidence interval, $\hat{I} = 10\,\% \pm 6\,\%$, is a rather wide interval and we wonder whether it is possible to reduce its range by using more than $M = 100$ testing samples. In case it were possible, we also wonder whether the range of $\hat{I}$ could be reduced to half the interval or even less; that is, whether it is possible to obtain $\hat{I} = 10\,\% \pm \hat{R}$ with $\hat{R} \le 3\,\%$. Regarding this issue, show the **CORRECT** statement:

A) In general, it is not possible to reduce the range of $\hat{I}$ because $\hat{I}$ does not significantly depend on M.
B) It is not possible to get a reduction of the range of $\hat{I}$ because we have considered that $\hat{p}$ will not significantly vary and so the range of $\hat{I}$ cannot vary either.
C) It is possible to reduce the range of $\hat{I}$ to half the interval or even less if we use at least twice as many testing samples ($M \ge 200$).
D) It is possible to reduce the range of $\hat{I}$ to half the interval or even less if we use at least four times as many testing samples ($M \ge 400$). $1.96 \cdot \sqrt{(0.1 \cdot 0.9)/M} \le 0.03 \rightarrow M \ge 385$

36 $\boxed{\text{B}}$ Let $\mathbf{x} = (x_1, \ldots, x_D)^t$, $D > 1$, be a D-dimensionobject that we want to classify in one among C classes. Which expression is *not* a minimum error (error risk) classifier?

A) $c(\mathbf{x}) = \arg\max_{c=1,\ldots,C}\; p(c)\, p(x_1 \mid c)\, p(x_2, \ldots, x_D \mid x_1, c)$
B) $c(\mathbf{x}) = \arg\max_{c=1,\ldots,C}\; \log p(x_1 \mid c) + \log p(x_2, \ldots, x_D \mid x_1, c)$
C) $c(\mathbf{x}) = \arg\max_{c=1,\ldots,C}\; p(c \mid x_1)\, p(x_2, \ldots, x_D \mid x_1, c)$
D) $c(\mathbf{x}) = \arg\max_{c=1,\ldots,C}\; \log p(x_1, c) + \log p(x_2, \ldots, x_D \mid x_1, c)$

37 $\boxed{\text{C}}$ For a three-class classification problem of objects of type $\mathbf{x}{=}(x_1, x_2)^t \in \{0,1\}^2$, we have the probability distributions shown in the table. Show the interval of the Bayes probability of error, ε^*:

A) $\varepsilon^* < 0.40$.
B) $0.40 \le \varepsilon^* < 0.45$.
C) $0.45 \le \varepsilon^* < 0.50$.
D) $0.50 \le \varepsilon^*$.

$\mathbf{x}$		$P(c \mid \mathbf{x})$			
x_1	x_2	$c{=}1$	$c{=}2$	$c{=}3$	$P(\mathbf{x})$
0	0	0.6	0.3	0.1	0
0	1	0.3	0.2	0.5	0.4
1	0	0.2	0.6	0.2	0.1
1	1	0.1	0.5	0.4	0.5

$\varepsilon^* = 0.49$

38 $\boxed{C}$ Suppose we have two boxes each containing 60 apples The first box contains 35 Gala apples and 25 Fuji apples. The second box contains 30 apples of each type. Let's assume we randomly pick a box, and then we randomly take an apple from the selected box. If the picked apple is Gala, which is the probability P that the apple is picked out of the first box? $P = 0.54$

A) $0/4 \leq P < 1/4$.

B) $1/4 \leq P < 2/4$.

C) $2/4 \leq P < 3/4$.

D) $3/4 \leq P \leq 4/4$.

39 $\boxed{B}$ The estimated probability of error of a classifier is 6 %. Which is the minimum number of testing samples, M, so that the 95 % confidence interval of this estimated probability of error is not higher than ± 1 %; that is, $I = [5\,\%, 7\,\%]$: $M = 2167$

A) $M < 2000$.

B) $2000 \leq M < 3500$.

C) $3500 \leq M < 5000$.

D) $M \geq 5000$.

40 $\boxed{B}$ In a problem of probabilistic reasoning corresponding to road trips, with the random variables: Climatology (C):{clear (CLE), cloudy (CLO), rainy (RAI)}; Luminosity (L):{day (DAY), night(NIG)}; Security (S):{secure (SEC), accident (ACC)}. The joint probability of the three random variables is provided by the table:

	DAY			NIG		
$P(s,l,c)$	CLE	CLO	RAI	CLE	CLO	RAI
SEC	0.27	0.17	0.06	0.16	0.13	0.07
ACC	0.01	0.02	0.02	0.03	0.02	0.04

The conditional probability $P(S = \text{ACC} \mid L = \text{NIG}, C = \text{CLE})$ is:

A) 0.190

B) 0.158

C) 0.030

D) 0.140

41 $\boxed{D}$ In a problem of probabilistic reasoning corresponding to road trips, with the random variables: Climatology (C):{clear (CLE), cloudy (CLO), rainy (RAI)}; Luminosity (L):{day (DAY), night(NIG)}; Security (S):{secure (SEC), accident (ACC)}. The joint probability of the three random variables is provided by the table:

	DAY			NIG		
$P(s,l,c)$	CLE	CLO	RAI	CLE	CLO	RAI
SEC	0.28	0.21	0.04	0.15	0.09	0.09
ACC	0.02	0.02	0.03	0.02	0.02	0.03

The conditional probability $P(S = \text{SEC} \mid L = \text{DAY}, C = \text{CLO})$ is:

A) 0.230

B) 0.210

C) 0.860

D) 0.913

42 $\boxed{D}$ The estimated probability of error of a classifier is 20 %. Which is the minimum number of testing samples, M, so that the 95 % confidence interval of this estimated probability of error is not higher than ± 1 %; that is, $I = [19\,\%, 21\,\%]$: $M = 6147$

A) $M < 2000$.

B) $2000 \leq M < 3500$.

C) $3500 \leq M < 5000$.

D) $M \geq 5000$.

43 $\boxed{C}$ For a three-class classification problem of objects of type $\mathbf{x}=(x_1, x_2)^t \in \{0,1\}^2$, we have the probability distributions shown in the table. Show the interval of the probability of error of the classifier $c(\mathbf{x})$ provided in the table, ε:

$\mathbf{x}$		$P(c \mid \mathbf{x})$			$P(\mathbf{x})$	$c(\mathbf{x})$
x_1	x_2	$c=1$	$c=2$	$c=3$		
0	0	0.2	0.1	0.7	0.2	2
0	1	0.4	0.3	0.3	0	1
1	0	0.3	0.4	0.3	0.4	3
1	1	0.4	0.4	0.2	0.4	1

$\varepsilon = 0.70$

A) $\varepsilon < 0.25$.

B) $0.25 \leq \varepsilon < 0.50$.

C) $0.50 \leq \varepsilon < 0.75$.

D) $0.75 \leq \varepsilon$.

44 $\boxed{C}$ Given the following joint frecuency distribution for the 3 random variables

A	0	0	0	0	1	1	1	1
B	0	0	1	1	0	0	1	1
C	0	1	0	1	0	1	0	1
N(A,B,C)	124	28	227	175	126	222	23	75

Which is the value of $P(A=1 \mid B=1, C=0)$?

A) 0.023

B) 0.250

C) 0.092

D) 0.446

45 $\boxed{C}$ For a three-class classification problem of objects of type $\mathbf{x}=(x_1, x_2)^t \in \{0,1\}^2$, we have the probability distributions shown in the table. Show the interval of the probability of error of the classifier $c(\mathbf{x})$ provided in the table, ε:

$\mathbf{x}$		$P(c \mid \mathbf{x})$			$P(\mathbf{x})$	$c(\mathbf{x})$
x_1	x_2	$c=1$	$c=2$	$c=3$		
0	0	0.2	0.3	0.5	0	1
0	1	0.3	0.3	0.4	0.4	1
1	0	0.2	0.5	0.3	0.5	2
1	1	0.3	0.6	0.1	0.1	1

$\varepsilon = 0.60$

A) $\varepsilon < 0.25$.

B) $0.25 \leq \varepsilon < 0.50$.

C) $0.50 \leq \varepsilon < 0.75$.

D) $0.75 \leq \varepsilon$.

46 $\boxed{C}$ Given the following joint frecuency distribution for the 3 random variables

A	0	0	0	0	1	1	1	1
B	0	0	1	1	0	0	1	1
C	0	1	0	1	0	1	0	1
N(A,B,C)	211	140	245	87	39	110	5	163

Which is the value of $P(A=1 \mid B=1, C=1)$?

A) 0.317

B) 0.163

C) 0.652

D) 0.250

Problems

1. (Exam November 26, 2012) In order to design a differential diagnosis between Flu and Cold, we use the histograms of body temperatures (fever) of a sample of patients who had these diseases. From the histograms, we get the following distribution of fever values in Celsius degrees:

$f(^oC)$	36	37	38	39	40
$P(F = f \mid D = \text{FLU})$	0.05	0.10	0.20	0.30	0.35
$P(F = f \mid D = \text{COLD})$	0.10	0.30	0.40	0.15	0.05

knowing that the relative occurrence of flu compared to cold is 30 % (i.e., $P(D = \text{FLU}) = 0.30$), calculate:

a) The conditional (posterior) probability that a patient has flu given he has a fever of 39^oC

b) The most probable diagnosis for this patient and the probability of error

c) The probabilities of diagnosis FLU and COLD $\forall f \in \{36, 37, 38, 39, 40\}$, and the minimum average probability of error ($P_\star(error)$) for a diagnosis system designed with the above observations.

Solution

a)

$$P(D = \text{FLU}) = 0.3; \qquad P(D = \text{COLD}) = 1 - 0.3 = 0.7$$

$$P(D = \text{FLU} \mid T = 39) = \frac{P(D = \text{FLU})P(T = 39 \mid D = \text{FLU})}{P(T = 39)}$$

$$P(T = 39) = P(D = \text{FLU})P(T = 39 \mid D = \text{FLU}) + P(D = \text{COLD})P(T = 39 \mid D = \text{COLD}) = 0.3 \cdot 0.3 + 0.7 \cdot 0.15 = 0.195$$

$$P(D = \text{FLU} \mid T = 39) = \frac{0.3 \cdot 0.3}{0.195} = 0.462$$

b)

$$P(D = \text{COLD} \mid T = 39) = \frac{0.7 \cdot 0.15}{0.195} = 0.538$$

Most probable diagnosis:

$$d^\star(T = 39) = \underset{d \in \{\text{FLU},\text{COLD}\}}{\arg\max} P(D = d \mid T = 39) = \text{COLD}$$

Probability that COLD is a wrong diagnosis for para $t = 39$:

$$P_\star(\text{error} \mid T = 39) = 1 - \text{máx}(P(D = \text{FLU} \mid T = 39), P(D = \text{COLD} \mid T = 39)) = 1 - \text{máx}(0.462, 0.538) = 0.462$$

c) Repeating the same calculations for $t \in \{36, 37, 38, 40\}$:

$t(^oC)$	36	37	38	39	40
$P(T = t)$	0.085	0.240	0.340	0.195	0.140
$P(D = \text{FLU} \mid T = t)$	0.176	0.125	0.176	0.462	0.750
$P(D = \text{COLD} \mid T = t)$	0.824	0.875	0.824	0.538	0.250
$P_\star(\text{error} \mid t)$	0.176	0.125	0.176	0.462	0.250

$$P_\star(\text{error}) = \sum_{t=36}^{40} P_\star(\text{error} \mid T = t)P(T = t) =$$

$$0.176 \cdot 0.085 + 0.125 \cdot 0.240 + 0.176 \cdot 0.340 + 0.462 \cdot 0.195 + 0.250 \cdot 0.140 = .230$$

2. **Exercise Irish Flowers, slide #22 of Chapter 1**

petal sizes in cm^2

$P(x \mid c)$	<1	1	2	3	4	5	6	7	8	9	10	>10
SETO	0.90	0.10	0	0	0	0	0	0	0	0	0	0
VERS	0	0	0	0.20	0.30	0.32	0.12	0.06	0	0	0	0
VIRG	0	0	0	0	0	0	0.08	0.12	0.24	0.14	0.20	0.22

Solution

Same prior probabilities

Assuming that the three classes have the same probability, calculate:

a) The conditional (posterior) probabilities $P(c \mid x)$, $c \in \{\textsc{Seto},\textsc{Vers},\textsc{Virg}\}$, for a flower whose petal size is $x = 7\,\text{cm}^2$

$$P(\text{SETO} \mid 7) = \frac{P(7 \mid \text{SETO})P(\text{SETO})}{P(7)} = \frac{0 * 0.33}{P(7)} = 0$$

$$P(\text{VERS} \mid 7) = \frac{P(7 \mid \text{VERS})P(\text{VERS})}{P(7)} = \frac{0.06 * 0.33}{P(7)} = \frac{0.06 * 0.33}{0.06} = \frac{0.02}{0.06} = 0.33$$

$$P(\text{VIRG} \mid 7) = \frac{P(7 \mid \text{VIRG})P(\text{VIRG})}{P(7)} = \frac{0.12 * 0.33}{P(7)} = \frac{0.12 * 0.33}{0.06} = \frac{0.04}{0.06} = 0.67$$

$P(7) = P(7 \mid \text{SETO}) * P(\text{SETO}) + P(7 \mid \text{VERS}) * P(\text{VERS}) + P(7 \mid \text{VIRG}) * P(\text{VIRG}) = 0 + 0.06 * 0.33 + 0.12 * 0.33 = 0.0198 + 0.0396 = 0.06$

b) The decision of optimal classification for this flower and the probability of taking a wrong decision.
Best decision for this flower is class Virginica. The probability of taking a wrong decision is 0.33 ($1 - \text{máx}_{d \in \mathcal{D}} P(d \mid x) = 1 - 0.67 = 0.33$). $P(error \mid 7) = 0.33$.

c) The best decision and probability of error for petals $1, 2, \ldots, 10$ cm^2
For petal = 1cm

$$P(\text{SETO} \mid 1) = \frac{P(1 \mid \text{SETO})P(\text{SETO})}{P(1)} = \frac{0.10 * 0.33}{0.033} = \frac{0.033}{0.033} = 1$$

$$P(\text{VERS} \mid 1) = \frac{P(1 \mid \text{VERS})P(\text{VERS})}{P(1)} = \frac{0 * 0.33}{P(1)} = \frac{0}{0.033} = 0$$

$$P(\text{VIRG} \mid 1) = \frac{P(1 \mid \text{VIRG})P(\text{VIRG})}{P(1)} = \frac{0 * 0.33}{P(1)} = \frac{0}{0.033} = 0$$

$P(1) = P(1 \mid \text{SETO}) * P(\text{SETO}) + P(1 \mid \text{VERS}) * P(\text{VERS}) + P(1 \mid \text{VIRG}) * P(\text{VIRG}) = 0.10 * 0.33 + 0 * 0.33 + 0 * 0.33 = 0.033$

This case is clear. The best decision is the only one possible: Setosa. So the probability of error is 0. $P(error \mid 1) = 0$.

For petal = 2cm

$$P(\text{SETO} \mid 2) = \frac{P(2 \mid \text{SETO})P(\text{SETO})}{P(2)} = \frac{0 * 0.33}{0} = \frac{0}{0} = 0$$

$$P(\text{VERS} \mid 2) = \frac{P(2 \mid \text{VERS})P(\text{VERS})}{P(2)} = \frac{0 * 0.33}{P(2)} = \frac{0}{0} = 0$$

$$P(\text{VIRG} \mid 2) = \frac{P(2 \mid \text{VIRG})P(\text{VIRG})}{P(2)} = \frac{0 * 0.33}{P(2)} = \frac{0}{0} = 0$$

$P(2) = P(2 \mid \text{SETO}) * P(\text{SETO}) + P(2 \mid \text{VERS}) * P(\text{VERS}) + P(2 \mid \text{VIRG}) * P(\text{VIRG}) = 0 * 0.33 + 0 * 0.33 + 0 * 0.33 = 0$

This case is clear. There is no best decision possible so the probability of error is 1. $P(error \mid 2) = 1$.

For petal = 3cm

$$P(\text{SETO} \mid 3) = \frac{P(3 \mid \text{SETO})P(\text{SETO})}{P(3)} = \frac{0 * 0.33}{0.033} = \frac{0}{0.067} = 0$$

$$P(\textsc{vers} \mid 3) = \frac{P(3 \mid \textsc{vers})P(\textsc{vers})}{P(3)} = \frac{0.20 * 0.33}{P(3)} = \frac{0.067}{0.067} = 1$$

$$P(\textsc{virg} \mid 3) = \frac{P(3 \mid \textsc{virg})P(\textsc{virg})}{P(3)} = \frac{0 * 0.33}{P(3)} = \frac{0}{0.067} = 0$$

$P(3) = P(3 \mid \textsc{seto}) * P(\textsc{seto}) + P(3 \mid \textsc{vers}) * P(\textsc{vers}) + P(3 \mid \textsc{virg}) * P(\textsc{virg}) = 0 * 0.33 + 0.20 * 0.33 + 0 * 0.33 = 0.067$

This case is clear. The best decision is the only one possible: Versicolor. So the probability of error is 0. $P(error \mid 3) = 0$.

For petal = 4cm

$$P(\textsc{seto} \mid 4) = \frac{P(4 \mid \textsc{seto})P(\textsc{seto})}{P(4)} = \frac{0 * 0.33}{0.033} = \frac{0}{0.01} = 0$$

$$P(\textsc{vers} \mid 4) = \frac{P(4 \mid \textsc{vers})P(\textsc{vers})}{P(4)} = \frac{0.30 * 0.33}{P(4)} = \frac{0.01}{0.01} = 1$$

$$P(\textsc{virg} \mid 4) = \frac{P(4 \mid \textsc{virg})P(\textsc{virg})}{P(4)} = \frac{0 * 0.33}{P(4)} = \frac{0}{0.06} = 0$$

$P(4) = P(4 \mid \textsc{seto}) * P(\textsc{seto}) + P(4 \mid \textsc{vers}) * P(\textsc{vers}) + P(4 \mid \textsc{virg}) * P(\textsc{virg}) = 0 * 0.33 + 0.30 * 0.33 + 0 * 0.33 = 0.01$

This case is clear. The best decision is the only one possible: Versicolor. So the probability of error is 0. $P(error \mid 4) = 0$.

For petal = 5cm

$$P(\textsc{seto} \mid 5) = \frac{P(5 \mid \textsc{seto})P(\textsc{seto})}{P(5)} = \frac{0 * 0.33}{0.033} = \frac{0}{0.11} = 0$$

$$P(\textsc{vers} \mid 5) = \frac{P(5 \mid \textsc{vers})P(\textsc{vers})}{P(5)} = \frac{0.33 * 0.33}{P(5)} = \frac{0.11}{0.11} = 1$$

$$P(\textsc{virg} \mid 5) = \frac{P(5 \mid \textsc{virg})P(\textsc{virg})}{P(5)} = \frac{0 * 0.33}{P(5)} = \frac{0}{0.11} = 0$$

$P(5) = P(5 \mid \textsc{seto}) * P(\textsc{seto}) + P(5 \mid \textsc{vers}) * P(\textsc{vers}) + P(5 \mid \textsc{virg}) * P(\textsc{virg}) = 0 * 0.33 + 0.32 * 0.33 + 0 * 0.33 = 0.11$

This case is clear. The best decision is the only one possible: Versicolor. So the probability of error is 0. $P(error \mid 5) = 0$.

For petal = 6cm

$$P(\textsc{seto} \mid 6) = \frac{P(6 \mid \textsc{seto})P(\textsc{seto})}{P(6)} = \frac{0 * 0.33}{P(6)} = \frac{0}{0.066} = 0$$

$$P(\textsc{vers} \mid 6) = \frac{P(6 \mid \textsc{vers})P(\textsc{vers})}{P(6)} = \frac{0.12 * 0.33}{P(6)} = \frac{0.0396}{0.066} = 0.60$$

$$P(\textsc{virg} \mid 6) = \frac{P(6 \mid \textsc{virg})P(\textsc{virg})}{P(6)} = \frac{0.08 * 0.33}{P(6)} = \frac{0.0264}{0.066} = 0.40$$

$P(6) = P(6 \mid \textsc{seto}) * P(\textsc{seto}) + P(6 \mid \textsc{vers}) * P(\textsc{vers}) + P(6 \mid \textsc{virg}) * P(\textsc{virg}) = 0 * 0.33 + 0.12 * 0.33 + 0.08 * 0.33 = 0.0396 + 0.0264 = 0.066$

The best decision is Versicolor. So the probability of error is 0.40. $P(error \mid 6) = 0.40$.

For petal = 8cm

$$P(\textsc{seto} \mid 8) = \frac{P(8 \mid \textsc{seto})P(\textsc{seto})}{P(8)} = \frac{0 * 0.33}{P(8)} = \frac{0}{0.08} = 0$$

$$P(\textsc{vers} \mid 8) = \frac{P(8 \mid \textsc{vers})P(\textsc{vers})}{P(8)} = \frac{0 * 0.33}{P(8)} = \frac{0}{0.08} = 0$$

$$P(\text{VIRG} \mid 8) = \frac{P(8 \mid \text{VIRG})P(\text{VIRG})}{P(8)} = \frac{0.24 * 0.33}{P(8)} = \frac{0.08}{0.08} = 1$$

$P(8) = P(8 \mid \text{SETO}) * P(\text{SETO}) + P(8 \mid \text{VERS}) * P(\text{VERS}) + P(8 \mid \text{VIRG}) * P(\text{VIRG}) = 0*0.33+0*0.33+0.24*0.33 = 0.0792$

This case is clear. The best decision is the only one possible: Virginica. So the probability of error is 0. $P(error \mid 8) = 0$.

For petal = 9cm

$$P(\text{SETO} \mid 9) = \frac{P(9 \mid \text{SETO})P(\text{SETO})}{P(9)} = \frac{0 * 0.33}{P(9)} = \frac{0}{0.08} = 0$$

$$P(\text{VERS} \mid 9) = \frac{P(9 \mid \text{VERS})P(\text{VERS})}{P(9)} = \frac{0 * 0.33}{P(9)} = \frac{0}{0.08} = 0$$

$$P(\text{VIRG} \mid 9) = \frac{P(9 \mid \text{VIRG})P(\text{VIRG})}{P(9)} = \frac{0.14 * 0.33}{P(9)} = \frac{0.0462}{0.0462} = 1$$

$P(9) = P(9 \mid \text{SETO}) * P(\text{SETO}) + P(9 \mid \text{VERS}) * P(\text{VERS}) + P(9 \mid \text{VIRG}) * P(\text{VIRG}) = 0*0.33+0*0.33+0.14*0.33 = 0.0462$

This case is clear. The best decision is the only one possible: Virginica. So the probability of error is 0. $P(error \mid 9) = 0$.

For petal = 10cm

$$P(\text{SETO} \mid 10) = \frac{P(10 \mid \text{SETO})P(\text{SETO})}{P(10)} = \frac{0 * 0.33}{P(10)} = \frac{0}{0.08} = 0$$

$$P(\text{VERS} \mid 10) = \frac{P(10 \mid \text{VERS})P(\text{VERS})}{P(10)} = \frac{0 * 0.33}{P(10)} = \frac{0}{0.08} = 0.$$

$$P(\text{VIRG} \mid 10) = \frac{P(10 \mid \text{VIRG})P(\text{VIRG})}{P(10)} = \frac{0.20 * 0.33}{P(10)} = \frac{0.066}{0.066} = 1$$

$P(10) = P(10 \mid \text{SETO}) * P(\text{SETO}) + P(10 \mid \text{VERS}) * P(\text{VERS}) + P(10 \mid \text{VIRG}) * P(\text{VIRG}) = 0*0.33+0*0.33+0.20*0.33 = 0.066$

This case is clear. The best decision is the only one possible: Virginica. So the probability of error is 0. $P(error \mid 10) = 0$.

d) The minimum probability of error for any iris flower; that is, $P_\star(\text{error})$
$P_\star(\text{error}) = \sum_x P(error \mid x)P(x) = P(error \mid < 1)P(< 1) + P(error \mid 1)P(1) + P(error \mid 2)P(2) + \ldots + P(error \mid 10)P(10) + P(error \mid > 10)P(> 10) = 0 + 0 + 0 + \ldots + 0.40 * 0.066 + 0.33 * 0.06 + 0 + 0 + 0 + 0 = 0.0264 + 0.0198 = 0.0462 \approx 0.05 = 5\,\%$ error

Section e): Different prior probabilities

Repeat the same calculations assuming that the prior probabilities are:
$P(\text{Seto}) = 0.3,\ P(\text{Vers}) = 0.5,\ P(\text{Virg}) = 0.2$

a) The conditional (posterior) probabilities $P(c \mid x)$, $c \in \{\text{Seto}, \text{Vers}, \text{Virg}\}$, for a flower whose petal size is $x = 7\,\text{cm}^2$

$$P(\text{SETO} \mid 7) = \frac{P(7 \mid \text{SETO})P(\text{SETO})}{P(7)} = \frac{0 * 0.30}{P(7)} = 0$$

$$P(\text{VERS} \mid 7) = \frac{P(7 \mid \text{VERS})P(\text{VERS})}{P(7)} = \frac{0.06 * 0.50}{P(7)} = \frac{0.06 * 0.50}{0.054} = \frac{0.03}{0.054} = 0.56$$

$$P(\textsc{virg} \mid 7) = \frac{P(7 \mid \textsc{virg})P(\textsc{virg})}{P(7)} = \frac{0.12 * 0.20}{P(7)} = \frac{0.12 * 0.20}{0.054} = \frac{0.024}{0.054} = 0.44$$

$P(7) = P(7 \mid \textsc{seto}) * P(\textsc{seto}) + P(7 \mid \textsc{vers}) * P(\textsc{vers}) + P(7 \mid \textsc{virg}) * P(\textsc{virg}) = 0 + 0.06 * 0.50 + 0.12 * 0.20 = 0.03 + 0.024 = 0.054$

b) The decision of optimal classification for this flower and the probability of taking a wrong decision.
Best decision for this flower is class Versicolor. The probability of taking a wrong decision is 0.44 ($1 - \text{máx}_{d \in \mathcal{D}} P(d \mid x) = 1 - 0.56 = 0.44$). $P(error \mid 7) = 0.44$.

c) For the exercise in section c) and d), we only need to compute the values for petal size 6 and 7, which are the petal sizes which condition the value of the minimum average probability of error (minimum probability of error for any iris flower) as the values for the rest of petal sizes is 0.

$$P(\textsc{seto} \mid 7) = \frac{P(7 \mid \textsc{seto})P(\textsc{seto})}{P(7)} = \frac{0 * 0.3}{P(7)} = 0$$

$$P(\textsc{vers} \mid 7) = \frac{P(7 \mid \textsc{vers})P(\textsc{vers})}{P(7)} = \frac{0.06 * 0.5}{P(7)} = \frac{0.06 * 0.5}{0.054} = \frac{0.03}{0.054} = 0.56$$

$$P(\textsc{virg} \mid 7) = \frac{P(7 \mid \textsc{virg})P(\textsc{virg})}{P(7)} = \frac{0.12 * 0.2}{P(7)} = \frac{0.12 * 0.2}{0.054} = \frac{0.024}{0.054} = 0.44$$

$P(7) = P(7 \mid \textsc{seto}) * P(\textsc{seto}) + P(7 \mid \textsc{vers}) * P(\textsc{vers}) + P(7 \mid \textsc{virg}) * P(\textsc{virg}) = 0 + 0.06 * 0.50 + 0.12 * 0.20 = 0.03 + 0.024 = 0.054$

$$P(\textsc{seto} \mid 6) = \frac{P(6 \mid \textsc{seto})P(\textsc{seto})}{P(6)} = \frac{0 * 0.30}{P(6)} = 0$$

$$P(\textsc{vers} \mid 6) = \frac{P(6 \mid \textsc{vers})P(\textsc{vers})}{P(6)} = \frac{0.12 * 0.50}{P(6)} = \frac{0.06}{0.076} = 0.79$$

$$P(\textsc{virg} \mid 6) = \frac{P(6 \mid \textsc{virg})P(\textsc{virg})}{P(6)} = \frac{0.08 * 0.20}{P(6)} = \frac{0.016}{0.076} = 0.21$$

$P(6) = P(6 \mid \textsc{seto}) * P(\textsc{seto}) + P(6 \mid \textsc{vers}) * P(\textsc{vers}) + P(6 \mid \textsc{virg}) * P(\textsc{virg}) = 0 + 0.12 * 0.50 + 0.08 * 0.20 = 0.06 + 0.016 = 0.076$

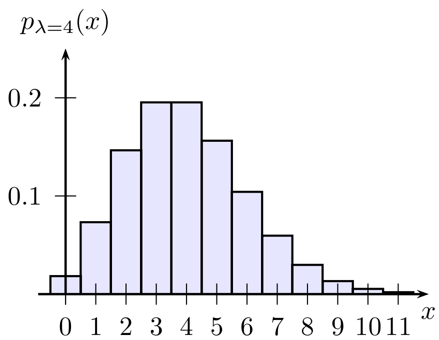
d) The minimum probability of error for any iris flower; that is, $P_\star(\text{error})$
$P_\star(\text{error}) = \sum_x P(error \mid x)P(x) = P(error \mid < 1)P(< 1) +$
$P(error \mid 1)P(1) + P(error \mid 2)P(2) + \ldots + P(error \mid 10)P(10) +$
$P(error \mid > 10)P(> 10) = 0+0+0+\ldots+0.21*0.076+0.44*0.054+0+0+0+0 = 0.016+0.024 = 0.04 \rightarrow 4\,\%$ error

3. (January 2020)

Let $\lambda \in \mathbb{R}^+$. We say that a random variable $x \in \{0, 1, 2, \ldots\}$ is Poisson(λ) if its probability mass function is:

$$p_\lambda(x) = \frac{\exp(-\lambda)\,\lambda^x}{x!}$$

The Poisson distribution is used to model the probability that one event occurs a certain number of times in a given context. The λ parameter can be interpreted as the average number of occurrences of the event. For instance, x can be the number of phone calls we receive in a day or the number of occurrences of a particular word in a document. The figure on the right shows $p_{\lambda=4}(x)$ for all $x \in \{0, 1, \ldots, 11\}$.



Let be a classification problem into C classes represented by means of a counter-type feature $x \in \{0, 1, 2, \ldots\}$. For every class c, we assume we know:

- The prior probability, $P(c)$.
- The conditional probability (mass) function, $P(x \mid c)$, which is Poisson(λ_c) with λ_c being known.

Answer the following questions:

a) (0.5 points) For the particular case: $C = 2$, $P(c = 1) = P(c = 2) = \frac{1}{2}$, $\lambda_1 = 1$, $\lambda_2 = 2$ and $x = 2$. Determine the unconditional probability of occurrence of $x = 2$, $P(x = 2)$.
b) (0.5 points) For the case given above, compute the posterior probability $P(c = 2 \mid x = 2)$, and the probability of error if $x = 2$ is classified in class $c = 2$.
c) (0.5 points) More generally, for any number of classes C and any value of prior probabilities, let's consider the case in which given $\tilde{\lambda} \in \mathbb{R}^+$, it holds $\lambda_c = \tilde{\lambda}$ for every c. In this case, there exists a class which is not dependent on x, c^*, under which every x can be classified with minimum probability of error. Determine which class is c^*.
d) (0.5 puntos) In the general case, prove that applying the Bayes classifier to this problem is equivalent to a classifier based on discriminant linear functions as follows (ln is the natural algorithm):

$$c^*(x) = \arg\max_c \; g_c(x) \quad \text{con} \quad g_c(x) = w_c\, x + w_{c0}, \quad w_c = \ln \lambda_c \quad \text{y} \quad w_{c0} = \ln p(c) - \lambda_c$$

Solution:

a) $P(x = 2 \mid c = 1) = \frac{1}{2e} = 0.1839 \qquad P(x = 2 \mid c = 2) = \frac{2}{e^2} = 0.2707.$
$P(x = 2) = 0.5 \cdot 0.1839 + 0.5 \cdot 0.2707 = 0.2273.$
b) $P(c = 2 \mid x = 2) = \frac{P(c=2) \cdot P(x=2|c=2)}{P(x=2)} = \frac{0.5 \cdot 0.2707}{0.2273} = 0.5955.$
$P(c \neq 2 \mid x = 2) = 1 - P(c = 2 \mid x = 2) = 0.4045.$
c) $c^*(x) = \arg\max_c P(c)\, P(x \mid c) = \arg\max_c P(c)\ \text{Poisson}(\lambda) = \arg\max_c P(c) = c^*.$
d)

$$\begin{aligned} c^*(x) &= \arg\max_c \; \ln P(c) + \ln P(x \mid c) \\ &= \arg\max_c \; \ln P(c) - \lambda_c + x \, \ln \lambda_c - \ln x! \\ &= \arg\max_c \; x \, \ln \lambda_c + (\ln P(c) - \lambda_c) \end{aligned}$$