Intelligent Systems
Exercises Block 2 Chapter 4
Clustering. Unsupervised learning: C-means algorithm

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

28 de octubre de 2016

1. Questions

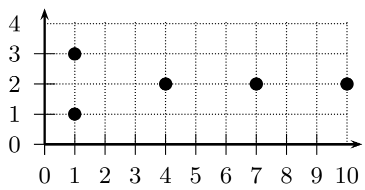
1 $\boxed{\text{C}}$ During the execution of the C-means algorithm, we obtain a partition which contains two clusters $X_1 = \{(0,0),(1,0),(2,1)\}$ and $X_2 = \{(0,1),(1,2),(2,2)\}$. Calculate the SSE (sum of squared errors) of this partition:

A) 8/3
B) 4/3
C) 16/3
D) 5/3

2 $\boxed{\text{B}}$ Regarding the SSE (sum of squared errors), show which of the following statements is TRUE:

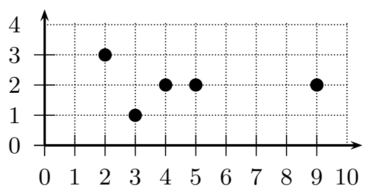
A) The Duda&Hart version of the C-means guarantees a global minimum of SSE
B) There is no polynomial cost algorithm that guarantees a global minimum of SSE
C) The Duda&Hart version of the C-means guarantees a null SSE (zero)
D) The "popular" version of the C-means guarantees a local minimum of SSE

3 $\boxed{\text{B}}$ The minimum value of SSE (sum of squared errors) to be able to group the data points of the figure on the right in two clusters is:



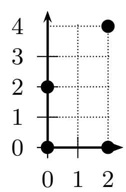
A) Lower than 10
B) Between 10 and 15
C) Between 15 and 20
D) Greater than 20

4 $\boxed{\text{B}}$ The minimum value of SSE (sum of squared errors) to be able to group the data points of the figure on the right in two clusters is:



A) Lower than 5.
B) Greater than 5 and lower than 10.
C) Greater than 10 and lower than 15.
D) Greater than 15.

5 $\boxed{\text{A}}$ The points in the figure on the right are grouped using the C-means algorithm, and after some running, the algorithm obtains the following partition $\Pi = \{X_1 = \{(0,0),(0,2)\}, X_2 = \{(2,0),(2,4)\}\}$, means $\mathbf{m}_1 = (0,1)$ and $\mathbf{m}_2 = (2,2)$, and SSE (sum of squared errors) $J = 10$. If the point $(2,0)$ is moved to another cluster, then:



A) The new value of SSE will be lower than 6.
B) The new value of SSE will be between 6 and 10.
C) The new value of SSE will be higher than 10.
D) It is not suitable to move the point because then the clusters would have uneven (unbalanced) sizes

6 $\boxed{\text{A}}$ Assume we have two classes A and B and that we have the following prototypes (samples) of each class: $A = \{(0,2),(1,1),(1,3),(2,2)\}$; and $B = \{(3,2),(3,3),(4,2),(4,3)\}$. Assume that these prototypes are two clusters that result from an unsupervised grouping process. The SSE value, J, of this partition would be:

A) $J \leq 6$

B) $6 < J \leq 8$
C) $8 < J \leq 10$
D) $J > 10$

7 $\boxed{\text{D}}$ The main difference between the Supervised Learning (SL) and Unsupervised Learning (UL) is:

A) in SL we know the correct class of the testing data and in UL we know the correct class of the training data.
B) in SL there is always a human operator who supervises the results so the system is merely used for assistance and in UL the whole process is automatically done
C) UL is an iterative process whereas SL is done at a time in a single step
D) in SL we know the correct class of all the data points and in UL we don't

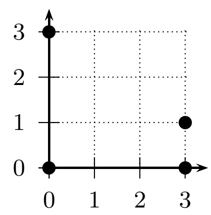
8 $\boxed{\text{B}}$ The C-means algorithm is a partitional clustering technique that we apply in speech recognition for ...

A) Transforming the voice (acoustic) signal into a parameterized signal
B) Designing *codebooks*
C) Training the Markov models
D) None of the above

9 $\boxed{\text{B}}$ Assume we have two classes A and B and that we have the following prototypes (samples) of each class: $A = \{(2,1),(1,2),(2,3),(3,2)\}$ and $B = \{(4,3),(5,3),(3,5),(6,5)\}$. Assume that these prototypes are two clusters that result from an unsupervised grouping process. The SSE value, J, of this partition would be:

A) $SSE < 4$
B) $SSE > 12$
C) $SSE = 11$
D) $4 < SSE < 10$

10 $\boxed{\text{C}}$ The points in the figure on the right are grouped by using the C-means algorithm, and after some running, the algorithm obtains the following partition $\Pi = \{X_1 = \{(0,0),(0,3),(3,0)\}, X_2 = \{(3,1)\}\}$. Let J' be the SSE value (sum of squared errors) of this partition, and let J be the SSE value of the partition that results from moving the point $(3,0)$ to another cluster. Therefore:

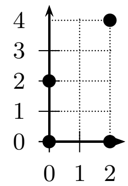


A) $J \geq J'$
B) $\frac{1}{2}J' \leq J < J'$
C) $\frac{1}{4}J' \leq J < \frac{1}{2}J'$
D) $J < \frac{1}{4}J'$

11 $\boxed{\text{C}}$ Regarding the unsupervised learning, which of the following statements is FALSE:

A) The goal of unsupervised learning is to group the data points in "natural" groupings
B) The SSE (Sum of Squared Errors) is a widely used measure to assess the quality of a partitional clustering
C) The C-means algorithm guarantees a global minimum of SSE
D) It is used, for instance, in Speech Recognition to represent an acoustic signal as a sequence of symbols associated to the "codewords"

12 $\boxed{\text{B}}$ The points of the figure on the right are grouped by using the C-means algorithm, and after some running, the algorithm obtains the following partition $\Pi = \{X_1 = \{(0,0),(0,2)\}, X_2 = \{(2,0),(2,4)\}\}$, means $\mathbf{m}_1 = (0,1)$ and $\mathbf{m}_2 = (2,2)$, and SSE (sum of squared errors) $J = 10$. If the point $(2,0)$ is moved to another cluster, then:



A) The new value of SSE will be lower than 5.
B) The new value of SSE will be between 5 and 7.
C) The new value of SSE will be higher than 7 but lower than 10
D) This point cannot be moved because otherwise one of the clusters would leave with only one data point.

13 $\boxed{\text{C}}$ Let $X = \{1, 3, 4.5\}$ be a set of three one-dimensional data points that we want to group into two clusters through a partitional clustering technique. In particular, we want to use the C-means algorithm and optimize the SSE value (sum of squared errors) but we have not decided yet whether to use the "popular" version or the *Duda and Hart (DH)* version. Let $\Pi^0 = \{X_1 = \{1,3\}, X_2 = 4.5\}$ be an initial partition which contains two clusters and $SSE_{\Pi^0} = J(\Pi_0) = 2$. Indicate which of the following statements is TRUE:

A) Both the "popular" and *DH* version will terminate without modifying the initial partition
B) The "popular" version will end with a better partition and the *DH* version will terminate with no modifications in the initial partition

C) The *DH* version will end with a better partition and the "popular" version will terminate with no modifications in the initial partition
D) Both versions will terminate with better partitions.

14 $\boxed{\text{A}}$ (Exam 18th January 2013) The criterion Sum of Square Errors (SSE) in partitional clustering is appropriate when the objects form:

A) Hyper-spherical clusters of similar size.
B) Hyper-spherical clusters of any size.
C) Elongated clusters of similar size.
D) Elongated clusters of any size.

15 $\boxed{\text{C}}$ (Exam 30th January 2013) We have three one-dimensional samples: $x_1 = 0$, $x_2 = 20$ y $x_3 = 35$, and the two-cluster partition $\Pi = \{X_1 = \{x_1, x_2\}, X_2 = \{x_3\}\}$. The sum of squared errors (SEE) of this partition is:

A) $J(\Pi) = 0$
B) $0 < J(\Pi) \leq 150$
C) $150 < J(\Pi) \leq 300$ $J(\Pi) = (x_1 - m_1)^2 + (x_2 - m_1)^2 + (x_3 - m_2)^2 = (0 - 10)^2 + (20 - 10)^2 + (35 - 35)^2 = 200$
D) $J(\Pi) > 300$

16 $\boxed{\text{B}}$ (Exam 30th January 2013) The application of the correct version of the K-means algorithm ("Duda and Hart") to the partition Π of the above question (question 15) yields the following resulting partition (Π^*): $\Delta J = \frac{n_2}{n_2+1}|x_2 - m_2|^2 - \frac{n_1}{n_1-1}|x_2 - m_1|^2$

A) $\Pi^* = \Pi$. $\Delta J = 0$
B) $\Pi^* = \{X_1 = \{x_1\}, X_2 = \{x_2, x_3\}\}$. $\Delta J = \frac{1}{2}|20 - 35|^2 - \frac{2}{1}|20 - 10|^2 = 112.5 - 200 = -87.5$
C) $\Pi^* = \{X_1 = \{x_2\}, X_2 = \{x_1, x_3\}\}$. $\Delta J = \frac{1}{2}|0 - 35|^2 - \frac{2}{1}|0 - 10|^2 = 612.5 - 200 = 412.5$
D) None of the above.

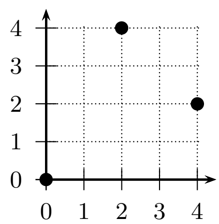
17 $\boxed{\text{D}}$ (Exam 15th January 2014) Which of the following statements about ***Clustering*** is *true*?:

A) The Perceptron algorithm is often used for labeled training samples
B) The Perceptron algorithm is often used for unlabeled training samples
C) The K-means algorithm is often used for labeled training samples
D) The K-means algorithm is often used for unlabeled training samples

18 $\boxed{\text{D}}$ (Exam 15th January 2014) The Sum of Square Errors (SSE) criterion is appropriate when the clusters are:

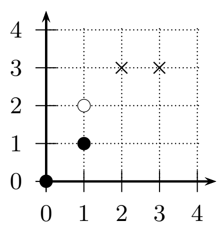
A) No elongated.
B) Elongated and of any size.
C) Elongated and of similar size.
D) None of the above.

19 $\boxed{\text{B}}$ (Exam 15th January 2014) The minimum value of the SSE (Sum of Square Errors) to group the samples on the right figure in two clusters is a value:



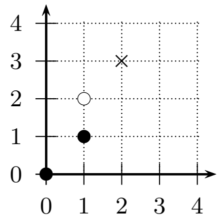
A) Between 0 and 3.
B) Between 3 and 6. $J = 4$
C) Between 6 and 9.
D) Greater than 9.

20 $\boxed{\text{B}}$ (Exam 15th January 2014) The figure on the right shows a partition of 5 bi-dimensional points in 3 clusters (represented with symbols $\bullet$, $\circ$ and $\times$). Consider all possible transfers of each point that is not in an unitary cluster. In terms of SSE (J):



A) No transfer improves J.
B) J can only improve transferring $(1,1)^t$ from cluster $\bullet$ to $\circ$.
C) J can only improve transferring $(2,3)^t$ from cluster $\times$ to $\circ$.
D) Both transfers in B) and C) improve J.

21 $\boxed{\text{C}}$ (Exam 30th January 2014) The figure on the right shows a partition of 4 bi-dimensional points in 3 clusters (represented with symbols $\bullet$, $\circ$ and $\times$). The Sum of Square Errors (SSE) for this partition is $J = 1$. If we apply the K-means algorithm (Duda and Hart version) to this partition:

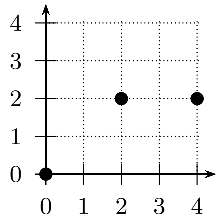


A) There will be no transfers between clusters.
B) A single point will be transferred, obtaining a partition with a J value between $\frac{2}{3}$ and 1.
C) A single point will be transferred, obtaining a partition with a J value between 0 and $\frac{2}{3}$.
J=0.5
D) There will be two transfers, obtaining a partition with $J = 0$.

22 $\boxed{\text{B}}$ (January 13, 2015) Which of the following assertions about *Clustering* is correct?:

A) The *K-means* algorithm is commonly used with *labeled* training samples
B) The *K-means* algorithm is commonly used with *unlabeled* training samples
C) The *Viterbi* algorithm is commonly used with *labeled* training samples
D) The *Viterbi* algorithm is commonly used with *unlabeled* training samples

23 $\boxed{\text{A}}$ (January 13, 2015) The minimum value of the SSE (Sum of Square Errors) to group the samples on the right figure in two clusters is a value:



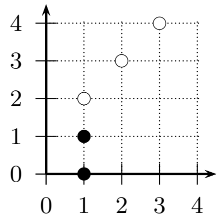
A) Between 0 and 3. $J = 2$

B) Between 3 and 6.

C) Between 6 and 9.

D) Greater than 9.

24 $\boxed{\text{C}}$ (January 13, 2015) The figure on the right shows a partition of 5 two-dimensional points in 2 clusters (represented with symbols $\bullet$ and $\circ$). Consider all possible cluster transfers of each point. The best transfer in terms of SSE (J) leads to an increment of SSE (ΔJ):



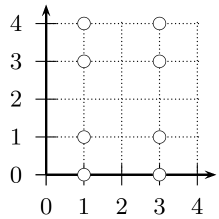
A) $\Delta J > 0$

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

C) $-1 \geq \Delta J > -2$ $\quad\Delta J = -1.5 \quad (J = 4.5 \rightarrow J = 3)$

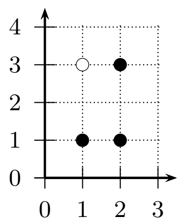
D) $-2 \geq \Delta J$

25 $\boxed{\text{B}}$ (January 26, 2015) The figure on the right shows 8 two-dimensional points. The minimum value of the Sum of Square Errors, J, to group the samples in two clusters is:



A) $0 \leq J \leq 7$
B) $7 < J \leq 14$ $\quad J = 10$
C) $14 < J \leq 21$
D) $21 < J$

26 $\boxed{\text{D}}$ (January 2016) 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) $-2 \geq \Delta J$ $\quad\Delta J = -\frac{21}{9} = -2.33 \quad (J = \frac{30}{9} \rightarrow J = 1)$

27 $\boxed{\text{B}}$ (January 2016) 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:

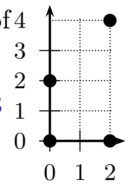
A) Both versions will get the same optimized partition

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

28 $\boxed{\text{A}}$ (January 2016) Consider the partition $\Pi = \{X_1 = \{(0,0)^t, (0,2)^t\}, X_2 = \{(2,0)^t, (2,4)^t\}\}$ for the points in the figure. The mean points of the clusters are $\mathbf{m}_1 = (0,1)^t$ and $\mathbf{m}_2 = (2,2)^t$. The Sum of Square Errors (SSE) of the partition is 10. If the point $(0,2)^t$ is transferred to cluster X_2, then:



A) The new SSE value will be >10. $\|(0,2)^t-(4/3,2)^t\|^2 + \|(2,0)^t-(4/3,2)^t\|^2 + \|(2,4)^t-(4/3,2)^t\|^2 = 32/3$

B) The new SSE value will be >8 and <10

C) The new SSE value will be >6 and <8

D) The new SSE value will be <6.

2. Problems

1. We have the following 5 two-dimensional vectors:

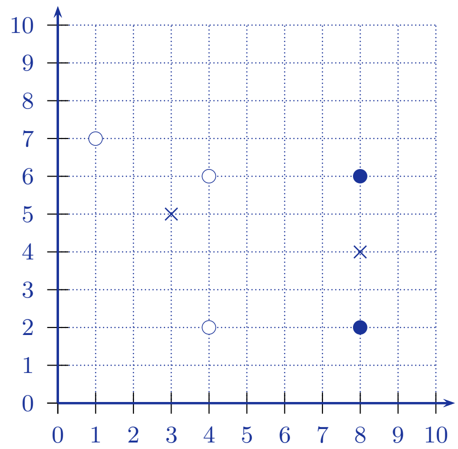
$$\mathbf{x}_1 = \begin{pmatrix}1\\7\end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix}4\\2\end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix}4\\6\end{pmatrix}, \quad \mathbf{x}_4 = \begin{pmatrix}8\\2\end{pmatrix} \quad \text{y} \quad \mathbf{x}_5 = \begin{pmatrix}8\\6\end{pmatrix}$$

We want to group the 5 vectors into two clusters by using unsupervised learning. Assuming we have the following initial partition:

$$\Pi = \{X_1 = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}, X_2 = \{\mathbf{x}_4, \mathbf{x}_5\}\}$$

trace the C-means algorithm and show one iteration of the main loop.

$$\mathbf{m}_1 = \frac{1}{3}(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3) = \begin{pmatrix}3\\5\end{pmatrix}$$
$$\mathbf{m}_2 = \frac{1}{2}(\mathbf{x}_4 + \mathbf{x}_5) = \begin{pmatrix}8\\4\end{pmatrix}$$



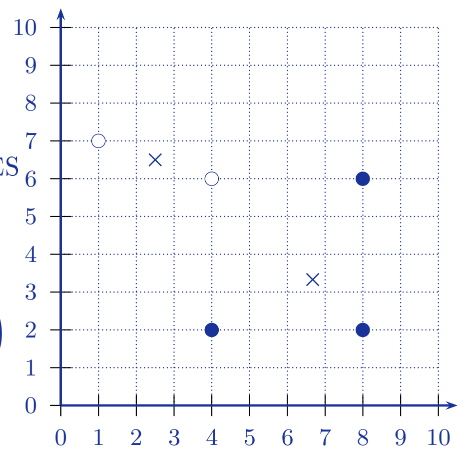
$$J_1 = \|\mathbf{x}_1 - \mathbf{m}_1\|^2 + \|\mathbf{x}_2 - \mathbf{m}_1\|^2 + \|\mathbf{x}_3 - \mathbf{m}_1\|^2 = 8 + 10 + 2 = 20$$
$$J_2 = \|\mathbf{x}_4 - \mathbf{m}_2\|^2 + \|\mathbf{x}_5 - \mathbf{m}_2\|^2 = 4 + 4 = 8$$
$$J = J_1 + J_2 = 28$$

If we transfer $\mathbf{x}_n \in X_i$ to X_j, then $\quad \Delta J = \dfrac{|X_j|}{|X_j|+1}\|\mathbf{x}_n - \mathbf{m}_j\|^2 - \dfrac{|X_i|}{|X_i|-1}\|\mathbf{x}_n - \mathbf{m}_i\|^2$

shall we transfer $\mathbf{x}_1$ from X_1 to X_2? : $\Delta J = \frac{2}{3} \cdot 58 - \frac{3}{2} \cdot 8 = \frac{80}{3} > 0 \Rightarrow$ NO

shall we transfer $\mathbf{x}_2$ from X_1 to X_2? : $\Delta J = \frac{2}{3} \cdot 20 - \frac{3}{2} \cdot 10 = -\frac{5}{3} < 0 \Rightarrow$ YES

$$\mathbf{m}_1 = \mathbf{m}_1 - \frac{\mathbf{x}_2 - \mathbf{m}_1}{|X_1| - 1} = \begin{pmatrix}5/2\\13/2\end{pmatrix} = \begin{pmatrix}2.5\\6.5\end{pmatrix}$$
$$\mathbf{m}_2 = \mathbf{m}_2 + \frac{\mathbf{x}_2 - \mathbf{m}_2}{|X_2| + 1} = \begin{pmatrix}20/3\\10/3\end{pmatrix} = \begin{pmatrix}6.67\\3.33\end{pmatrix}$$
$$J = J + \Delta J = \frac{79}{3} = 26.33$$



$$\text{shall we transfer } \mathbf{x}_3 \text{ from } X_1 \text{ to } X_2\text{? } : \quad \Delta J = \frac{3}{4}\cdot\frac{128}{9} - \frac{2}{1}\cdot\frac{10}{4} = \frac{17}{3} = 5.67 > 0 \;\Rightarrow\; \text{NO}$$

$$\text{shall we transfer } \mathbf{x}_4 \text{ from } X_2 \text{ to } X_1\text{? } : \quad \Delta J = \frac{2}{3}\cdot\frac{151}{2} - \frac{3}{2}\cdot\frac{32}{9} = \frac{805}{16} = 50.31 > 0 \;\Rightarrow\; \text{NO}$$

$$\text{shall we transfer } \mathbf{x}_5 \text{ from } X_2 \text{ to } X_1\text{? } : \quad \Delta J = \frac{2}{3}\cdot\frac{61}{2} - \frac{3}{2}\cdot\frac{80}{9} = 7 > 0 \;\Rightarrow\; \text{NO}$$

(THE SOLUTION TO THE PROBLEM ENDS HERE). The algorithm continues as follows:

$$\text{shall we transfer } \mathbf{x}_1 \text{ from } X_1 \text{ to } X_2\text{? } : \quad \Delta J = \frac{3}{4}\cdot\frac{410}{9} - \frac{2}{1}\cdot\frac{5}{2} = \frac{175}{6} = 29.17 > 0 \;\Rightarrow\; \text{NO}$$

$$\text{shall we transfer } \mathbf{x}_2 \text{ from } X_2 \text{ to } X_1\text{? } : \quad \Delta J = \frac{2}{3}\cdot\frac{45}{2} - \frac{3}{2}\cdot\frac{80}{9} = \frac{5}{3} = 1.67 > 0 \;\Rightarrow\; \text{NO}$$

No more transfers will be done so we don't need to continue. The optimized partition is:

$$\Pi = \{X_1 = \{\mathbf{x}_1, \mathbf{x}_3\}, X_2 = \{\mathbf{x}_2, \mathbf{x}_4, \mathbf{x}_5\}\}$$