Guía de ejercicios Capítulo III

Análisis de Agrupamientos

# Objetivos de la Unidad

|  |  |  |
| --- | --- | --- |
| **#** | **Descripción** | **Preguntas** |
| 1. | Establecer diferencias entre agrupamientos jerárquicos y no jerárquicos. |  |
| 2. | Comprender los conceptos de similaridad en espacios n-dimensionales como un concepto de distancia. |  |
| 3. | Comprender la estructuración de un agrupamiento jerárquico. |  |
| 4. | Cuantificar las medidas de similaridad y su aplicación a la agrupación. |  |
| 5. | Comprender los algoritmos básicos de los agrupamientos. |  |
| 6. | Comprender las medidas de calidad para evaluar agrupamientos. |  |

# Preguntas

P1. Agrupamiento de genes

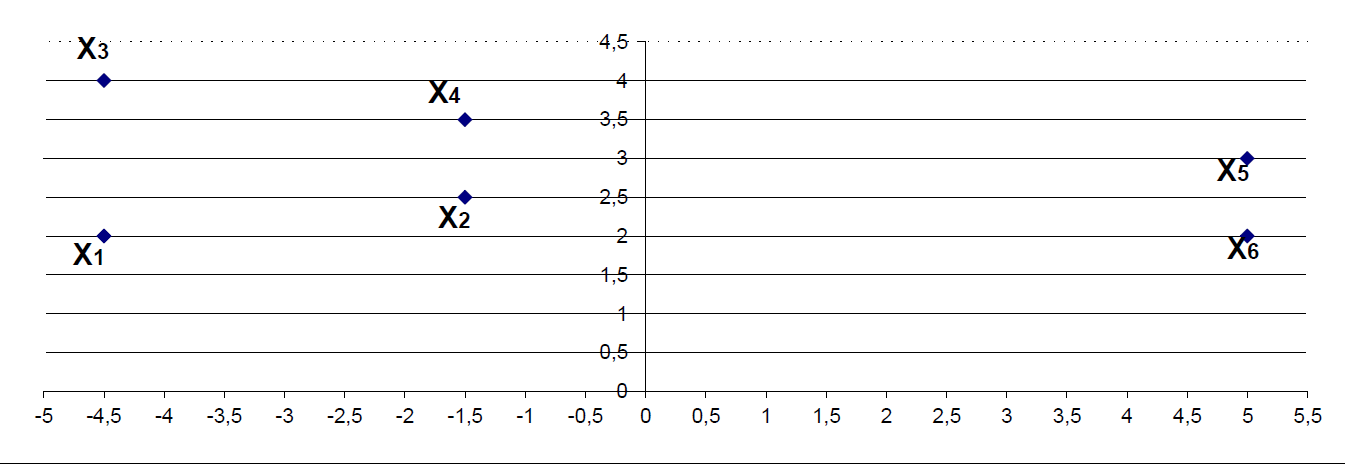
En una gran variedad de problemas de bioinformática, se requiere determinar grupos de genes que intervienen en una determinada enfermedad. Los genes están formados por un alfabeto básico que contiene 4 letras (bases) A, G, C, T. Muchas veces con un trozo de gen (aproximadamente 8 bases “letras”) es posible caracterizar un gen. Para la siguiente tabla determine:

1. Un método para medir distancia entre genes.
2. Construya una matriz de distancia entre los genes.
3. Determine el dendograma que caracteriza este conjunto de genes.
4. Determine 2, 3 o 4 grupos según corresponda.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gen 1 | G | A | T | A | C | A | T | T |
| Gen 2 | G | A | T | A | C | A | T | A |
| Gen 3 | G | A | T | A | C | T | A | C |
| Gen 4 | C | T | A | A | G | G | G | G |
| Gen 5 | C | T | C | A | G | G | G | G |
| Gen 6 | G | A | T | T | T | C | C | G |
| Gen 7 | G | A | T | T | A | C | C | G |

1. Para la siguiente distribución de casos determine la matriz de distancias y el árbol

Jerárquico (dendrograma) en forma aproximada.



1. Para confeccionar un catálogo de obras de arte se requería agrupar las diferentes obras. Se reunió a un grupo de 16 expertos que clasificaron las obras por pares en forma binaria (‘1’=parecidas, ‘0’=diferentes). La matriz de similaridades se compone de las sumas de las votaciones ‘1’ de los expertos. Use un método de agrupamiento jerárquico para obtener el dendrograma resultante y entregue dos conjuntos que cada uno de ellos representen obras similares.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | X1 | X2 | X3 | X4 | X5 | X6 |
| X1 | 16 | 15 | 15 | 3 | 15 | 6 |
| X2 |  | 16 | 1 | 15 | 3 | 15 |
| X3 |  |  | 16 | 15 | 3 | 15 |
| X4 |  |  |  | 16 | 15 | 6 |
| X5 |  |  |  |  | 16 | 15 |
| X6 |  |  |  |  |  | 16 |

1. Aproxime a valores enteros los centros de los centroides y dibuje el dendrograma resultante al aplicar un agrupamiento jerárquico a la siguiente matriz.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | X1 | X2 | X3 | X4 | X5 | X6 |
| X1 | 0 | 1,2 | 2,3 | 8,2 | 8,3 | 8,4 |
| X2 |  | 0 | 2,1 | 8,1 | 8,2 | 8,2 |
| X3 |  |  | 0 | 8,3 | 8,2 | 8,2 |
| X4 |  |  |  | 0 | 3,3 | 3,2 |
| X5 |  |  |  |  | 0 | 2,3 |
| X6 |  |  |  |  |  | 0 |

# Respuestas

P1. XXX

Texto.