**Exercícios: agrupamento aglomerativo**

1. Utilize os métodos de agrupamento aglomerativo de distância mínima e distância máxima para agrupar os dados descritos pela seguinte matriz de proximidade. Mostre os dendrogramas.



1. Aplique o método de *clustering* hierárquico aglomerativo aos dados da tabela abaixo para obter uma árvore de grupos, tendo em conta as noções de distância a seguir indicadas, e represente a árvore obtida através de um dendrograma. Analise esse dendrograma e indique qual o nível (nº de grupos) da hierarquia que lhe parece constituir uma partição mais natural dos dados. Compare os resultados.



* 1. Distância entre dados: *Manhattan*; distância entre grupos: mínima.
  2. Distância entre dados: *Manhattan*; distância entre grupos: máxima.

1. Utilize os métodos de agrupamento aglomerativo de distância mínima, distância máxima e distância média para agrupar os oito exemplos seguintes: A1 = (2, 10), A2 = (2, 5), A3 = (8, 4), A4 = (5, 8), A5 = (7, 5), A6 = (6, 4), A7 = (1, 2), A8 = (4, 9). Mostre os dendrogramas.
2. Desenhe à mão a árvore de aglomeração para a amostra seguinte constituída por 10 pontos numa só dimensão, (-22, -20, -3, 1, 2, 4, 16, 17, 19, 20). Utilize o método hierárquico de aglomeração “*single linkage*”.
   1. Baseando-se na árvore de grupos que obteve, porque é que é mais natural a divisão dos dados em 3 agrupamentos?
3. Considere as seguintes espécies de hominídeos, cada um descrito por duas características: massa cerebral e massa corporal (H. = Homo; A. = Australopithecus, E = early e L = late). Agrupe estas espécies, aplicando o algoritmo hierárquico aglomerativo de distância mínima, considerando distâncias de *Manhattan* e os valores dos atributos normalizados.
   1. Apresente o dendrograma correspondente.
   2. Indique, justificando, em quantos grupos optaria por dividir estas espécies.



1. Aplique o método de agrupamento hierárquico aglomerativo aos dados da tabela abaixo para obter uma árvore de grupos e represente a árvore obtida através de um dendrograma. Tenha em conta a distância euclidiana entre pontos de dados e as duas distâncias entre os grupos: distância mínima e distância máxima.

