Departamento de Matemática

Multivariate Analysis
Mater in Eng. and Data Science & Master in Mathematics and Applications

2nd Test 1st Semester – 2020/2021
Duration: 1.5 hours   17/12/2020 – 18:30

Please justify conveniently your answers

Group I   **9.0 points**

1. Let $\boldsymbol{x} = (x_1, \ldots, x_p)^t$ and $\boldsymbol{y} = (y_1, \ldots, y_p)^t$ represents the p continuous measurements characteri- (3.0)
   zing two different objects, where for $\boldsymbol{x}, \boldsymbol{y} \in (I\!\!R^+)^p$,

   $$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\frac{1}{p}\sum_{i=1}^{p}\left(\frac{x_i - y_i}{x_i + y_i}\right)^2} \quad \text{and} \quad d(\boldsymbol{0}, \boldsymbol{0}) = 0.$$

   Prove that the previous coefficient is a dissimilarity measure.

2. Show that the distance $d_{k(ij)}$ (between cluster k and the cluster formed by merging cluster i and (2.0)
   j) used by average-linkage verifies:

   $$d_{k(ij)} = \frac{n_i}{n_i + n_j}d_{ki} + \frac{n_j}{n_i + n_j}d_{kj},$$

   where d_{ij} is the dissimilarity between i-th and j-th cluster and n_i is the number of objects belonging
   to the i-th cluster.

3. The pairwise dissimilarities between four objects are as follows:

   $$\boldsymbol{D} = \begin{bmatrix} 0 & 1 & 11 & 5 \\ & 0 & 2 & 3 \\ & & 0 & 4 \\ & & & 0 \end{bmatrix}.$$

   Use average-linkage cluster analysis on the dissimilarity matrix above, and draw the associated (4.0)
   dendrogram.
   Suggestion: Use the result stated in Question 2.

Group II   **11.0 points**

1. An observation x comes from one of two populations with prior probabilities $P(Y = 0) = P(Y = 1)$
   and probability mass functions:

   $$P(X = x|Y = j) = \begin{cases} \frac{\lambda_j^x e^{-\lambda_j}}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise}, \end{cases}$$

   with $\lambda_1 > \lambda_0 > 0$, $j = 0, 1$, and $X|Y = j \sim \text{Poisson}(\lambda_j)$.

   (a) Obtain the classification rule that minimizes the total probability of misclassification. (4.0)
   (b) Assuming $\lambda_0 = 1$, and $\lambda_1 = 4$, obtain:
       i. The recall of class 1. (2.5)
       ii. The precision of class 1. (2.5)
       iii. The F_1 measure of class 1. (1.0)
   (c) Comment on the accuracy of the classifier. (1.0)