	Ahsanullah University of Science and Technology		
	Department of Computer Science and Engineering		
	SET A, Class Test #3, Fall 2023		
	Course Code: CSE 4261	Course Title: Data Analytics	
	Time: 20 Minutes	Date: 01/6/2024	Full Marks: 10
ID:		Name:	

1. Interpret the following Autocorrelation and Partial Autocorrelation Functions (Figure 1 and Figure 2) and decide the optimum "p" value for an Autoregression (AR) Model for Figure 1 and the optimum "q" value for a Moving Average (MA) Model for Figure 2.

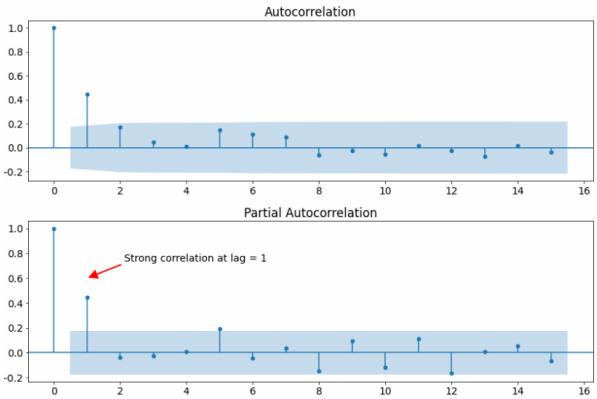


Figure 1: ACF and PACF figures for the AR model

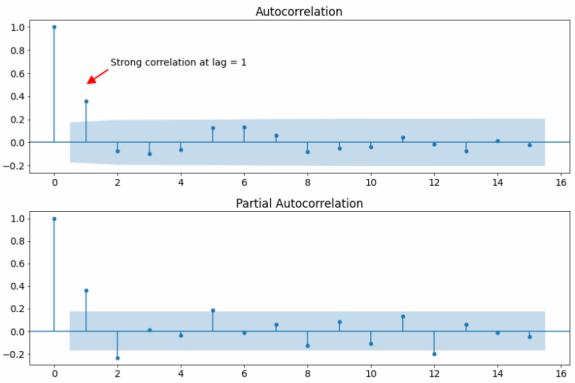
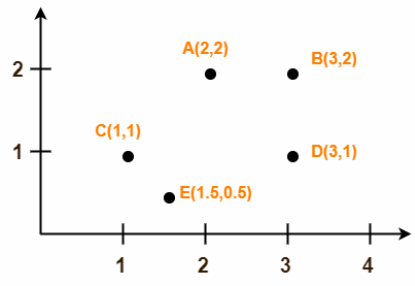


Figure 2: ACF and PACF figures for the MA model

2. Use Forgy's and Kmean's algorithms to create two clusters on the following datasets. The distance function between two points a = (x1, y1) and b = (x2, y2) is defined as-D (a, b) = |x2 – x1| + |y2 – y1|. Use the Dunn Index to compare the performance of the two clustering algorithms.



	Department of Computer Science and Engineering		
	SET B, Class Test #3, Fall 2023		
	Course Code: CSE 4261	Course Title: Data Analytics	
	Time: 20 Minutes	Date: 01/6/2024	Full Marks: 10
ID:		Name:	

1. Interpret the following Autocorrelation and Partial Autocorrelation Functions (Figure 1 and Figure 2) and decide the optimum "p" value for an Autoregression (AR) Model for Figure 1 and the optimum "q" value for a Moving Average (MA) Model for Figure 2.

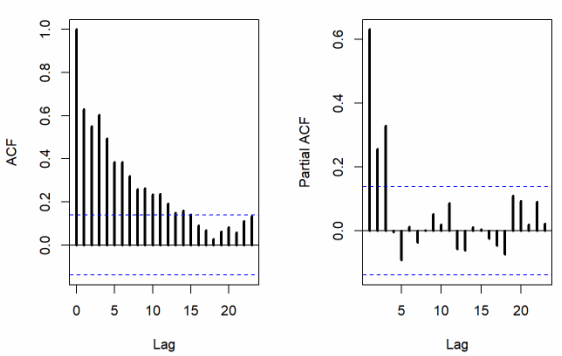


Figure 1: ACF and PACF figures for the AR model

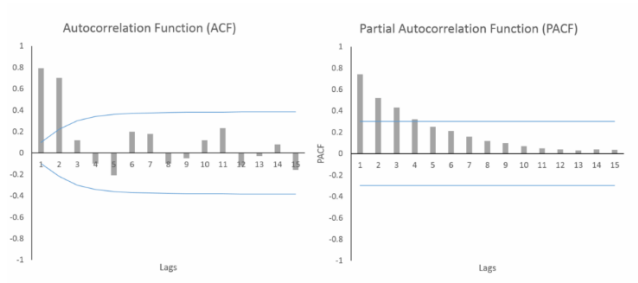


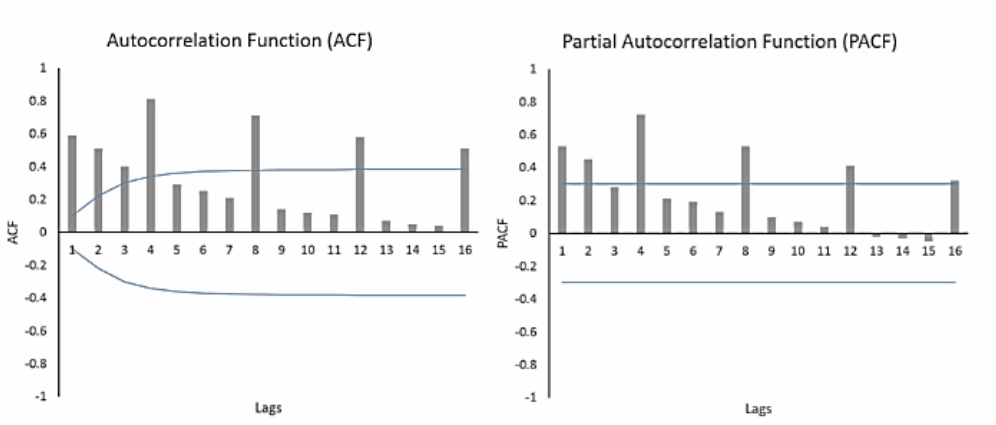
Figure 2: ACF and PACF figures for the MA model

2. The equation below assesses the partition quality in Conceptual-based Clustering (COBWEB). Explain the significance of the i, j, and k variables, as well as the terminology used to formulate the equation.

$$\sum_{k=1}^{n} P(C_k) \sum_{i} \sum_{j} P(A_i = V_{ij} | C_k)^2$$

	Ahsanullah University of Science and Technology		
	Department of Computer Science and Engineering		
	SET C, Class Test #3, Fall 2023		
	Course Code: CSE 4261	Course Title: Data Analytics	
	Time: 20 Minutes	Date: 01/6/2024	Full Marks: 10
ID:		Name:	

1. Determine the appropriate SARIMA model for the following autocorrelation and partial autocorrelation functions. You must express the grounds for your decision.



2. Investigate the equation and find the use of this equation in the Fuzzy C-mean algorithm for clustering a dataset. Explain the significance of the i, t, j, c, and m variables, as well as the terminology used to formulate the equation.

$$A_i^{(t+1)}(x_k) = \left[\sum_{j=1}^{c} \left(\frac{\left\| x_k - v_i^{(t)} \right\|^2}{\left\| x_k - v_j^{(t)} \right\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}$$

	Ahsanullah University of Science and Technology		
	Department of Computer Science and Engineering		
	SET D, Class Test #3, Fall 2023		
	Course Code: CSE 4261	Course Title: Data Analytics	
	Time: 20 Minutes	Date: 01/6/2024	Full Marks: 10
ID:		Name:	

1. Interpret the following Autocorrelation and Partial Autocorrelation Functions (Figure 1 and Figure 2) and decide the optimum "p" value for a Seasonal Autoregression (AR) Model for Figure 1 and the optimum "q" value for a Seasonal Moving Average (MA) Model for Figure 2.

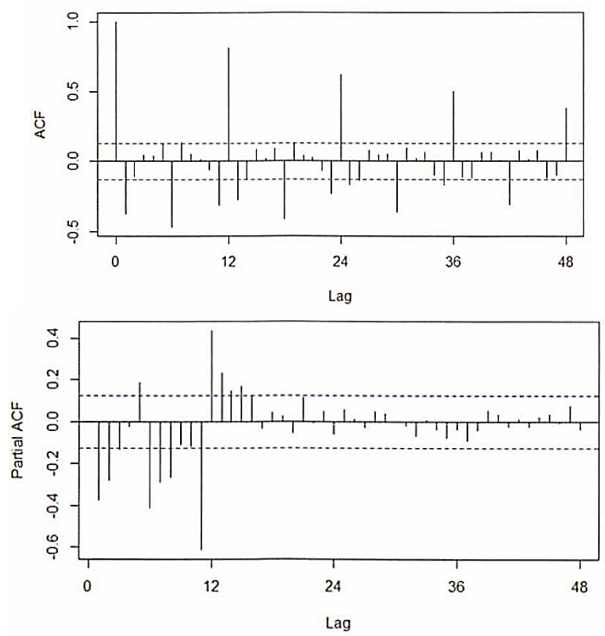


Figure 1: ACF and PACF figures for the AR model

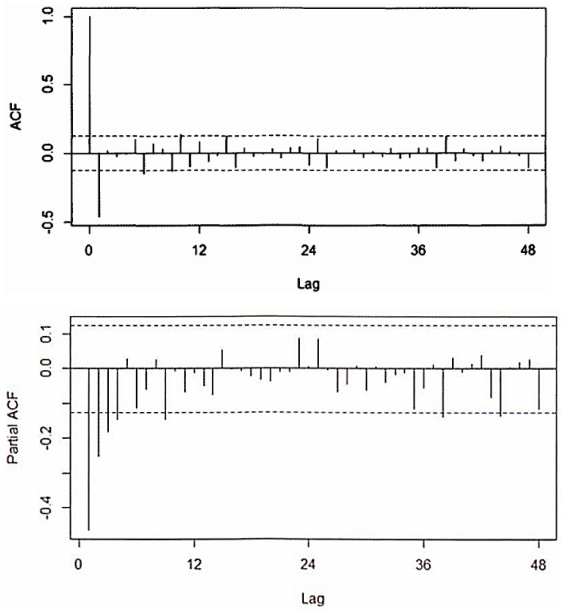


Figure 2: ACF and PACF figures for the MA model

2. Use Forgy's and Kmean's algorithms to create two clusters on the following datapoints (1,2), (2,2), (4,4.5), and (2,1.5). The distance function between two points a = (x1, y1) and b = (x2, y2) is defined as Euclidian distance. Use the Davies-Bouldin Index to compare the performance of the two clustering algorithms.