

Assignment-2 Applied Data Science

1. For the given data, compute two clusters using K-means algorithm for clustering where initial cluster centres are (1.0, 1.0) and (5.0, 7.0). Execute for two iterations.

Record Number	A	B
R1	1.0	1.0
R2	1.5	2.0
R3	3.0	4.0
R4	5.0	7.0
R5	3.5	5.0
R6	4.5	5.0
R7	3.5	4.5

- 2.
3. Given the points A(3, 7), B(4, 6), C(5, 5), D(6, 4), E(7, 3), F(6, 2), G(7, 2) and H(8, 4), Find the core points and outliers using DBSCAN. Take Eps = 2.5 and MinPts = 3.
4. What are the limitations of the ARIMA time series model when dealing with nonlinear and non-stationary data? Provide examples of data types or scenarios where ARIMA may not perform well and suggest alternative approaches for such cases.
5. Discuss the advantages of using fuzzy time series forecasting over crisp time series methods in scenarios with high uncertainty or noise. Provide a real-world example of data that could benefit from a fuzzy approach and explain why traditional methods might struggle with it.
6. Discuss cases in which content-based recommendations will not perform as well as ratings-based collaborative filtering.
7. What is a recommender system? Explain the Content based and Collaborative filtering based recommender systems by mentioning their limitations.
8. Consider the following ratings table between five items and 7 users:

Item-ID	User 1	User 2	User 3	User 4	User 5	User 6	User 7
1	5	6	7	4	3	?	5
2	4	?	3	?	5	4	?
3	?	3	4	1	1	?	4
4	7	4	3	6	?	4	6
5	1	?	3	2	2	5	2

- (a) Predict the values of unspecified ratings of user 2 using user-based collaborative filtering algorithms. Use the Pearson correlation with mean-centering.
- (b) Predict the values of unspecified ratings of user 2 using item-based collaborative filtering algorithms. Use the adjusted cosine similarity.

9. Suppose that you had a recommender system that could predict raw ratings. How would you use it to design a top-k recommender system? Discuss the computational complexity of such a system in terms of the number of applications of the base prediction algorithm. Under what circumstances would such an approach become impractical?
10. What is clustering? Differentiate between partitional clustering and hierarchical clustering.
11. Explain the K-Means clustering Algorithm? How the choice of initial cluster centres affect the convergence of K-Means clustering?
12. Mathematically explain ARIMA(p,d,q), SARIMA (p,d,q) (P,D,Q)_m and simple exponential smoothing models.
13. Given multiple ARIMA(p,d,q) models for a time series. How will you choose the right model based on AIC, AICc and BIC criterion?
14. Given the following parameters for a time series: 10,20, 25, 30. Forecast the next value of the time series using Simple Exponential Smoothing (SES), Linear Trend and Damped Trend models.

term	SES	Linear trend	Damped trend
α	1.00	1.00	1.00
β^*		0.30	0.40
ϕ			0.98
ℓ_0	10.28	10.05	10.04
b_0		0.22	0.25

15. Explain the steps followed in time series forecasting using machine learning models. Take an example and explain step by step.
16. Explain the steps followed in fuzzy time series forecasting methods using machine learning techniques. Take an example and explain step by step.
17. Mathematically explain the Holt and Winters additive method for handling time series with seasonal patterns.
18. Mathematically explain the AR (2) model. What are the parameters of the model?
19. Mathematically explain the MA(q) model. How it can be realized using ARIMA(p,d,q).
20. Write a short note on SARIMA model.