



CONCURRENT EVALUATION FOR BUSINESS ANALYTICS SPECIALISATION

III SEM / ADVANCE STATISTIC METHOD USING R (304)

Faculty Name: Prof. Ujval More

Important Instructions:

1. The subject is evaluated on the basis of three components

Component Number	Component	Marks	Submission Date
1	Written Home Assignment	50	10 th Feb 2023
2	Caselet	50	13 th Feb 2023
3	Term Paper	50	

2. Assignments to be submitted in person.

3. Student Name, contact number, email-id, Specialization, Component must be clearly mentioned

COMPONENT 1: Written Home Assignment

Q1. What is Business Hypothesis? Write five hypotheses (alternate hypothesis and null hypothesis)

Q2. What is logistic regression? Explain it with suitable example by using any statistical tool like Excel, SPSS, Python, R

Q3. What do you mean by significance? Testing a correlation for significance by using sample data of

Table 1

Sr no	City	Temperature (0°)	Wind speed (m/s)
1	Pune	32	7.2
2	Nasik	28	5.6
3	Sangali	31	4.5
4	Jalgaon	35	7.1
5	Bhusawal	34	4.5
6	Kolhapur	29	6.1
7	Barmati	27	7.2

Q4. What do you mean by cross tabulation? What are the different function used in Rfor calculating cross tabulation.

Q5. Explain the statistical tools and describe their applications in multiple businessdomains and scenarios

Q6. What is Linear regression ? Calculate linear regression of the above data (use Table-1)

Q7. Explain Dimension Reduction ? what are the different stages Dimension Reduction Techniques..

Q8. Explain Logistic Regression. What are the steps to perform logistic regression is R .Q9.

What is Log likelihood ratio test ? Explain it with suitable example.

Q10. What is Time Series objects in R ? Explain Decomposition of Time Series.

COMPONENT 1: Caselet

Caselet-1

In the internet era, prediction of customer behavior is a very valuable insight, since it helps a marketer to analyse its products' value and send updates for selling its products. The online market depends on the history of its customers. Devising new strategies for markets and attracting customers to stores and trying to convert the incoming traffic into sales profitably are all vital to the financial health of retailers. Every retailer uses different strategies to increase store traffic and convert traffic into profits. They invest in prime real estate with desirable properties such as high foot-traffic of their targeted customer segments, customer populations, customer convenience and visibility. Once they determine a location, retailers drive store traffic in a variety of ways such as spending on advertising, offering loss-leader about the products with various discounts or conducting various promotional events in local markets, such as offering discounts at various levels or price deductions. Whenever costumers visit a store, retailers try to convert the customers profitably through several means. They ensure that the right product is available at the right place, at the right time and at the right price. They invest in store labour to ensure that costumers experience a good and competitively priced shopping service that would encourage them to purchase and return to the store in future as well. Such relationships are critical to retailers for the following reasons. First, they get to know the feedback of other stores and requirements of the customers. Financial data of the local customers can be calculated using time series data. Decision tree is very important for this type of problem as we can calculate the risk factors in the local market and understands the needs of the customers from their previous behavior. This is also known as learning the cognitive behavior of the customer. Let us take the example of iphone 7

that was launched recently. This brand also uses time-series analysis for understanding the behavior of their customers by means of data gathered from the earlier models like iPhone 6 and iPhone 6s. How the customer used these earlier models and what features they look for in competitive products provides important insights for product development. Decision tree is very useful for gathering information about new market values as these depend on the time series that comes from historical data. Using such data, we can analyse information from new products as well. We can analyse customer behavior in conjunction with their financial status and give them best discounts for their needs. If we analyse historical data, many products have failed badly because they were not able to understand the requirements of the market at that time. So, to play it safe, every company nowadays tries to understand the market and its needs as per the market values, thus, creating a decision tree from the time-series data is an essential task for them. Decision trees can help in reducing errors by means of information gain from the parent to the child. Tree based induction in ID3 helps to generate a recommendation engine. Such an engine is a powerful tool to understand the needs of the market and help companies choose profitable markets. Decision trees have many features that are very helpful to retailers and companies for offering discounts by comparing the information gain and loss in the market. This is also done by understanding the behavior of the customer with regards to the new product and older products - iPhone 6 and 6s being pertinent examples here because after launching iPhone 7 and 7s the prices of iPhone 6 and 6s were reduced by 20k in the Indian Market. Using decision tree and its properties in data mining, we can increase the profits for retailers and help companies convert customer traffic into profits.

Questions

1. What are the features of decision trees? (10)
2. Define properties in data mining, by which we can increase the profits for retailers? (10)
3. Analyze the case and your opinion about this case.

Caselet 2

The term, 'recommender systems' is widely used nowadays. Recommender systems are composed of very simple algorithms that aim to provide the most relevant and accurate information to users by sorting/filtering useful information from very large databases. Recommendation engines discover data patterns from a given dataset by learning the consumers' information and then producing outcomes that correlate to their needs and interests. In addition, recommendation engines narrow down the risk that could become a complex decision to just a few recommendations search. Big data supports recommendations at an unimaginable level these days. Recommendation engines work mainly in one of the following two ways, viz., either they rely on the properties of items with their bread crumbs that a user likes, which are analysed to determine what else the user may like, or they rely on the likes and dislikes of other users, which the recommendation engine uses to compute a similarity index between users and recommends items to them accordingly. It is also possible to combine both these methods to build a highly-advanced recommendation engine. The main goal is to achieve the recommended collective information of users for the items that might interest customers. These systems have access to user-centric information with profile attributes, such as demographics and product descriptions. They differ in the way they interact while analyzing the data to develop affinity values between users and items, which can be used to identify well-matched pairs. A collaborative filtering system is used for matching and analyzing historical interaction alone, while content-based filtering is used for profiling-based attributes. Let us see how we can implement a recommendation engine with a collaborative memory-based recommendation engine. However, before that we must first understand the logic behind such a system. To this engine, each item and each user is nothing but an identifier or token element. Let us take the example of Netflix. Please note that we will not take any other attribute of a movie, such as cast, director, genre, etc., into consideration while generating recommendations for users. The similarity between two users is represented by using a decimal number between -1.0 and 1.0. We will call this number, the similarity index. The possibility of a user liking a movie will be represented by using another decimal number between -1.0 and 1.0. Now that we have

modeled the world around this system using simple terms, we can unleash a handful of elegant mathematical equations to define the relationship between these identifiers and numbers. In our recommendation algorithm, we will maintain a number of sets, which should represent a member of supersets with all users and identifiers. Each user will have two sets, viz., a set of movies the user likes and a set of movies the user dislikes, Each movie will also have two sets associated with it, viz., a set of users who liked the movie and a set of users who disliked the movie. During the performance where recommendations start to generate, a number of sets will be produced, mostly unions or intersections of the other sets. We will also have ordered lists of suggestions and similar users for each user. Similarly, like movies we can use the following recommendations. Personalised Product Information E-commerce Sites Such engines help in understanding customers' preferences on the basis of their visit on the website. They show the customers the most relevant recommendation-type products as per their needs or there likes in real time. Recommendation improve as the cognitive learning improves with regression about each visitor each time. Website Personalisation This is used by many organizations to calculate revenue on the basis of the number of hits from visitors. It increases their sales and targets new customers through segmentation into different cluster. It also allows getting in touch by message-centric methods. Real-time Notifications This is used by e-commerce for letting their customers know about the new top selling brands and available discounts. Such engines help brands build trust among their customers and create a sense of presence and urgency while showing real-time notification of shoppers' activities on their website.

Questions

1. Which filtering system is used for matching and analyzing historical interaction alone and define?
2. Why content-based filtering is used for profiling-based attributes?
3. What are the different number of sets generated?

COMPONENT 3: TERM PAPER

MBA-II / SEM-III / (2022-23)

END SEMESTER EVALUATION

Subject: ADVANCE STATISTIC METHOD USING R
Faculty Name: Prof. Ujjval More

Marks: 50]

Instructions to the candidates:

1. All questions are compulsory
2. All questions carry 10 marks
3. Figures to the right indicate full marks

Q1) Answer the following questions (Any Five)

[5×2=10]

- a) Define NULL and Alternate hypothesis.
- b) Define statistical modeling
- c) What is adjusted R^2 in regression analysis?
- d) Explain Unlist () function.
- e) Explain aov () function.
- f) What is logistic regression.
- g) Define Predictive analytics.
- h) How many predictor variables must be used in multiple regression?

Q2) Answer the following questions. (Any two).

[2×5=10]

- a) Compare Poisson distribution and binomial distribution.
- b) Describe test procedure for testing significance of Correlation coefficient.
- c) Explain Z test of hypothesis testing. Write the syntax and explain in detail.

Q3) Solve any one of the following. **[1×10=10]**

- a) Consider the employee salary database and perform all types of descriptive analysis the data with the help of R programming code.
- b) Explain multiple regression with its two application.

Q4) Solve any one of the following. **[1×10=10]**

- a) Explain dimension reduction techniques with example.
- b) Discuss techniques of performance evaluation of logistic regression model.

Q5) Solve any one of the following. **[1×10=10]**

- a) Explain the concept of Time Series Discuss how time series is used in business forecasting.
- b) Describe linear Discriminant analysis (LDA). Write a brief outline of R code for the same.