**Machine Learning and Statistics**

**Q1.** Consider a 2-Class classification task in the 3-Dimensional space, where the two classes C1 and C2 are modeled by Gaussian distribution with mean m1 = [0,0,0] and m2 = [0.5,0.5,0.5] respectively. The Covariance matrix is for both the distribution is

S= [[0.8,0.01,0.01],

[0.01,0.2,0.01],

[0.01,0.01,0.2]]

Given point x=[0.1,0.5,0.1] classify x

1.Using Euclidean distance.

2. Mahalanobis distance

Also, comment on results.

Step 1. Use the function euclidean\_classifier

x=[0.1 0.5 0.1]; m1=[0 0 0]; m2=[0.5 0.5 0.5]

m=[m1 m2]

z=euclidean\_classifier(m,x)

The answer is z = 1, that is, the point is classified to the C1 class.

Step 2. Use the function mahalanobis\_classifier

x=[0.1 0.5 0.1]; m1=[0 0 0]; m2=[0.5 0.5 0.5]; m=[m1 m2];

S=[0.8 0.01 0.01;0.01 0.2 0.01; 0.01 0.01 0.2];

z=mahalanobis\_classifier(m,S,x);

This time, the answer is z = 2, meaning the point is classified to the second class. For this case, the optimal Bayesian classifier is realized by the Mahalanobis distance classifier. The point is assigned to class C2 in spite of the fact that it lies closer to m1 according to the Euclidean norm.

**Q2.** Generate 50 2-Dimensional feature vectors from a Gaussian distribution N(**μ**,S)

Where

**μ** = [ 2 , -2] , S = [[0.9 , 0.2],

[0.2 , 0.3]]

Let *X* be the resulting matrix having the features vector as columns. Compute the maximum likelihood(ML) estimate of mean value **μ,** and the covariance matrix, S, of N(**μ**,S) and comment on the resulting estimates.

To generate X,

randn('seed',0);

μ= [2 -2];

S = [0.9 0.2; 0.2 .3];

X = mvnrnd(m,S,50);

To compute the ML estimates of μ and S

[μ\_hat, S\_hat]=Gaussian\_ML\_estimate(X);

The results are

μ\_hat= [2.0495, −1.9418] , S\_hat = [0.8082, 0.0885; 0.0885 ,0.2298]

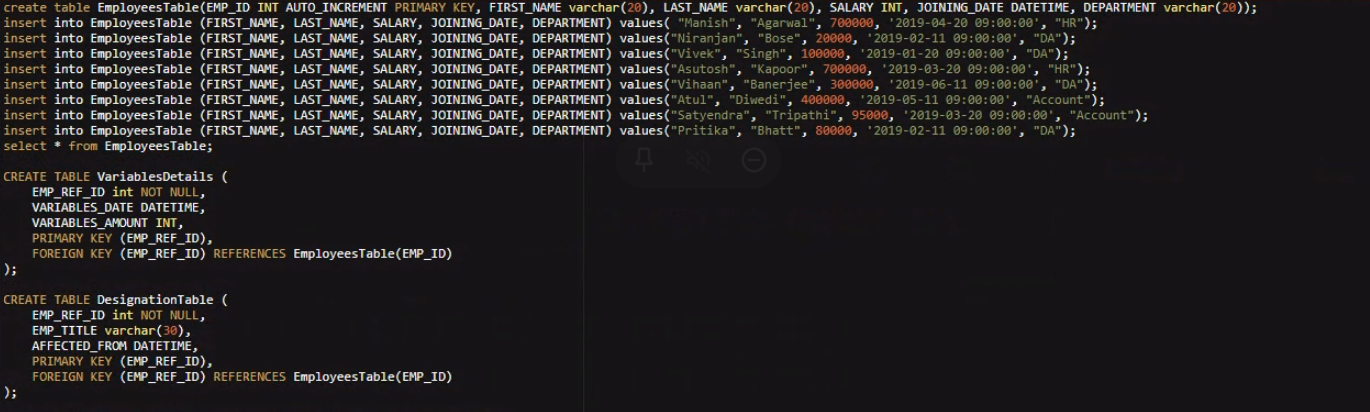
It can be observed that the estimates that define the corresponding Gaussian distribution, although close to the true values of the parameters, cannot be trusted as good estimates. This is due to the fact that 50 points are not enough to result in reliable estimates. The returned values depend on the initialization of the random generator (involved in function mvnrnd), so there is a slight deviation among experiments.

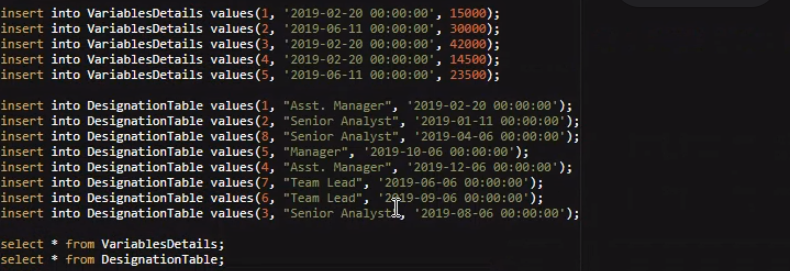
**SQL**

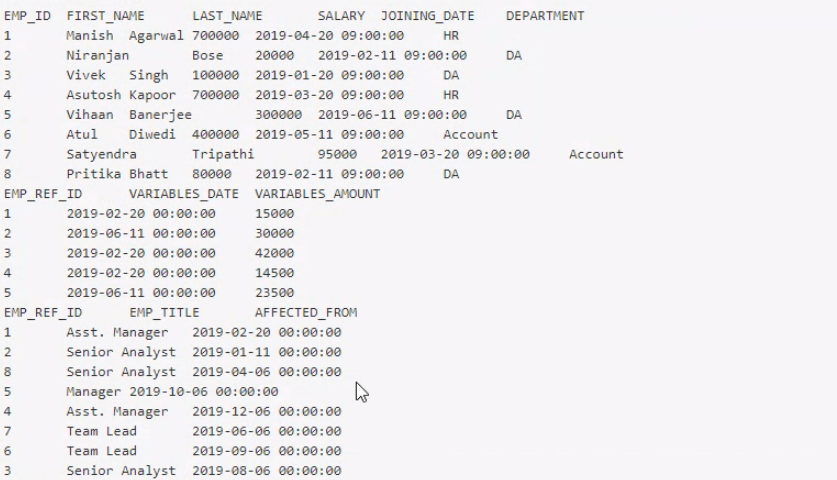
1.Write a SQL query to create these tables in your database and insert the data into these tables with the following characteristics:

a. Add the primary key “Emp\_ID” to the Employees Table. Also, mention what are the constraints used in SQL.

b. Add foreign key “EMP\_REF\_ID” in Variables Details and Designation Table that references “Emp\_ID” in Employees Table



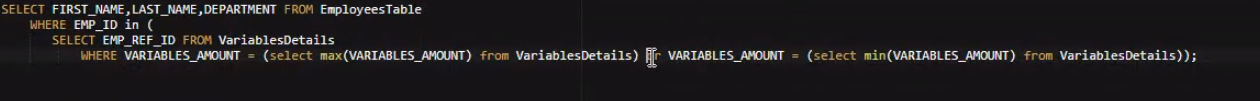




2. What are the four different types of joins? Give examples of each by performing the joins on the Employees table and Designation Table.

* INNER JOIN: Returns records that have matching values in both tables
* LEFT (OUTER) JOIN: Returns all records from the left table, and the matched records from the right table
* RIGHT (OUTER) JOIN: Returns all records from the right table, and the matched records from the left table
* FULL (OUTER) JOIN: Returns all records when there is a match in either left or right table

a. Write a query to get the employee details(full name and department) who received the highest and the least variables





c. What is cross join? Write a query to give an example of the same by performing it on the Employees table and Designation table.

d. What are the clauses used with Select statements and what are the orders of it? Write a query to get the employee details who got their designations updated in the second half of the year 2019(July to December), sorted by the “variables\_amount” (highest to lowest) where the department name of the Employee has the letter ‘A’ in it.

C. What is cross join? Write a query to give an example of the same by performing it on the Employees table and Designation table.

Ans. Cross joins are used to return every combination of rows from two tables. It is also known as cartesian product.

D. What are the clauses used with Select statements and what are the orders of it? Write a query to get the employee details who got their designations updated in the second half of the year 2019(July to December), sorted by the “variables\_amount” (highest to lowest) where the department name of the Employee has the letter ‘A’ in it.

Ans. The SELECT statement has the following clauses: SELECT , FROM, WHERE, GROUP BY, HAVING, ORDER BY, OFFSET, FETCH FIRST, UNION, INTERSECT, EXCEPT, WITH

3. What is a Cursor? Write a query/queries to use the cursor to store the Employees Name( full name) for the HR department into a variable called ‘emp\_names’

Ans. Cursor is a temporary memory that is used to store Database Tables.

A. What is Normalization and explain different forms of normalization with examples. (preferable with the above tables)

Ans. Normalization is a process to organize the data to reduce redundancy from the table.

First normal form(1NF) : No column should have multi valued attribute

Second normal form (2NF): Table must be in 1NF and all the non prime attributes should be fully functional dependent on candidate key (No partial dependency)

Third normal form(3NF) : It should be in 2NF and No transitive dependency that column should not depend on the columns which are not a part of primary key.

BCNF(3.5NF) : It should be in 3NF, for every [functional dependency](https://beginnersbook.com/2015/04/functional-dependency-in-dbms/) X->Y, X should be the super key of the table.

B. What is the stored procedure? Write a stored procedure to call the query that you have written for Q2.a

Ans. A stored procedure is a prepared SQL code that you can save, so the code can be reused over and over again.