

Worksheet 3

Machine Learning Assessment 3

1. Which of the following is an application of clustering?

Ans. (d) All of the above

2. On which data type, we cannot perform cluster analysis?

Ans. (d) None

3. Netflix's movie recommendation system uses

Ans. (d) All of the above

4. The final output of Hierarchical clustering is

Ans. (a) The number of cluster centroids

5. Which of the step is not required for K-means clustering?

Ans. (d) None

6. Which is the following is wrong?

Ans. (d) None

7. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?

Ans. (d) All of the above

8. Which of the following are true? i. Clustering analysis is negatively affected by multicollinearity of features ii. Clustering analysis is negatively affected by heteroscedasticity

Ans. (a) i only

9. In the figure above, if you draw a horizontal line on y-axis for $y=2$. What will be the number of clusters formed?

Ans. (d) 5

10. For which of the following tasks might clustering be a suitable approach?

Ans. (b) Given a database of information about your users, automatically group them into different market segments

11. Given, six points with the following attributes:

Ans. (b)

12. Ans. (a)

13. What is the importance of clustering?

Ans. Data analysis and data mining applications rely heavily on clustering. It is the challenge of arranging a collection of things so that those in the same group are more comparable to those in other groups. A decent clustering algorithm can identify clusters regardless of their form.

14. How can I improve my clustering performance?

Ans. There are various approaches of evaluating the performance of the clustering algorithm. First, compare it to something that is known to operate successfully. Then compare the outcomes. Second, time the algorithms and compare the times of the two. If there are two sets of good answers, one may examine how the solution's quality evolves over time. Third, test the algorithms on several instances of a problem. One is not too difficult, one is medium, and one is quite difficult. Finally, employing evolution to optimise the parameters of your clustering algorithms might put it to the test under pressure and push you in the right way for improvement.

SQL Assessment 3

```
{
  "cells": [
    {
      "cell_type": "code",
      "execution_count": 1,
      "id": "324985ab",
      "metadata": {},
      "outputs": [],
      "source": [
        "import sqlite3"
      ]
    },
    {
      "cell_type": "code",
      "execution_count": 3,
      "id": "db032fbe",
      "metadata": {},
      "outputs": [],
      "source": [
        "db = sqlite3.connect('customer_database')"
      ]
    },
    {
      "cell_type": "markdown",
      "id": "8c258ca8",
```

```

"metadata": {},
"source": [
  "Q1. Write SQL query to create table Customers."
]
},
{
  "cell_type": "code",
  "execution_count": 7,
  "id": "7f56d6fc",
  "metadata": {},
  "outputs": [
    {
      "data": {
        "text/plain": [
          "<sqlite3.Cursor at 0x22ef74638f0>"
        ]
      },
      "execution_count": 7,
      "metadata": {},
      "output_type": "execute_result"
    }
  ],
  "source": [
    "db.execute('create table customer (customernumber int primary key, customer name text not null, contactlastname text, contactfirstname text, phone int not null, adressline text, city text, state text, postalcode int, country text, salesrepeployeeenumbebr int, creditlimit int)')"
  ]
},
{
  "cell_type": "code",
  "id": "4ab8ba59",
  "metadata": {},
  "source": [
    "Q2. Write SQL query to create table Orders."
  ]
},
{
  "cell_type": "code",
  "execution_count": 8,
  "id": "3868b779",
  "metadata": {},
  "outputs": [
    {
      "data": {

```

```

    "text/plain": [
      "<sqlite3.Cursor at 0x22ef7463ea0>"
    ]
  },
  "execution_count": 8,
  "metadata": {},
  "output_type": "execute_result"
}
],
"source": [
  "db.execute('create table orders (ordernumber int primary key, orderdate int not null, requireddate int, shippeddate int, status text, comments text, customernumber int not null)')"
],
{
  "cell_type": "markdown",
  "id": "1b7e0d42",
  "metadata": {},
  "source": [
    "Q3. Write SQL query to show all the columns data from the Orders Table."
  ]
},
{
  "cell_type": "code",
  "execution_count": 9,
  "id": "ec033def",
  "metadata": {},
  "outputs": [
    {
      "data": {
        "text/plain": [
          "[]"
        ]
      },
      "execution_count": 9,
      "metadata": {},
      "output_type": "execute_result"
    }
  ],
  "source": [
    "result = db.execute('select * from orders')\n",
    "result.fetchall()"
  ]
},

```

```

{
  "cell_type": "markdown",
  "id": "8977d279",
  "metadata": {},
  "source": [
    "Q4. Write SQL query to show all the comments from the OrdersTable."
  ]
},
{
  "cell_type": "code",
  "execution_count": 11,
  "id": "1938a418",
  "metadata": {},
  "outputs": [
    {
      "data": {
        "text/plain": [
          "[]"
        ]
      },
      "execution_count": 11,
      "metadata": {},
      "output_type": "execute_result"
    }
  ],
  "source": [
    "result = db.execute('select comments from orders')\n",
    "result.fetchall()"
  ]
},
{
  "cell_type": "markdown",
  "id": "45bde17",
  "metadata": {},
  "source": [
    "Q5. Write a SQL query to show orderDate and Total number of orders placed on that date, from Orderstable."
  ]
},
{
  "cell_type": "code",
  "execution_count": 12,
  "id": "2b5db5a9",
  "metadata": {},

```

```

"outputs": [
  {
    "data": {
      "text/plain": [
        "[(None, None)]"
      ]
    },
    "execution_count": 12,
    "metadata": {},
    "output_type": "execute_result"
  }
],
"source": [
  "result = db.execute('select orderdate, sum(ordernumber) from orders')\n",
  "result.fetchall()"
],
{
  "cell_type": "markdown",
  "id": "8c0d3971",
  "metadata": {},
  "source": [
    "Q6. Write a SQL query to show employeeNumber, lastName, firstName of all the employees from employees table."
  ]
},
{
  "cell_type": "code",
  "execution_count": 13,
  "id": "f08b0e13",
  "metadata": {},
  "outputs": [
    {
      "data": {
        "text/plain": [
          "<sqlite3.Cursor at 0x22ef748d570>"
        ]
      },
      "execution_count": 13,
      "metadata": {},
      "output_type": "execute_result"
    }
  ],
  "source": [

```

```
"db.execute('create table employees (employeenumber int primary key, lastname text,
firstname text, extension int not null, email text, officecoad int, reportto text, jobtitle text)')
```

```
]
},
{
  "cell_type": "code",
  "execution_count": 14,
  "id": "56f81b03",
  "metadata": {},
  "outputs": [
    {
      "data": {
        "text/plain": [
          "[]"
        ]
      },
      "execution_count": 14,
      "metadata": {},
      "output_type": "execute_result"
    }
  ],
  "source": [
    "result = db.execute('select employeenumber, lastname, firstname from employees')\n",
    "result.fetchall()"
  ]
},
{
  "cell_type": "markdown",
  "id": "c15da413",
  "metadata": {},
  "source": [
    "Q7. Write a SQL query to show all orderNumber, customerName of the person who
    placed the respective order."
  ]
}
```

```
]
},
{
  "cell_type": "code",
  "execution_count": 16,
  "id": "a3e28172",
  "metadata": {},
  "outputs": [
    {
      "data": {
        "text/plain": [

```

```

    "[]"
  ]
},
"execution_count": 16,
"metadata": {},
"output_type": "execute_result"
}
],
"source": [
  "result = db.execute('select ordernumber from orders ')\n",
  "result1 = db.execute('select customer name from customer')\n",
  "result.fetchall()\n",
  "result1.fetchall()"
]
},
{
  "cell_type": "markdown",
  "id": "2a81fcff",
  "metadata": {},
  "source": [
    "Q8. Write a SQL query to show name of all the customers in one column and salerepemployee name in another column. "
  ]
},
{
  "cell_type": "raw",
  "id": "d0961344",
  "metadata": {},
  "source": [
    "result = db.execute('select customer name from customer')\n",
    "result 1 = db.execute('select salerepemployee from emplpoyee')\n",
    "result.fetchall()\n",
    "result1.fetchall()"
  ]
},
{
  "cell_type": "markdown",
  "id": "b9bde056",
  "metadata": {},
  "source": []
}
],
"metadata": {
  "kernel_spec": {

```



```

"display_name": "Python 3 (ipykernel)",
"language": "python",
"name": "python3"
},
"language_info": {
"codemirror_mode": {
"name": "ipython",
"version": 3
},
"file_extension": ".py",
"mimetype": "text/x-python",
"name": "python",
"nbconvert_exporter": "python",
"pygments_lexer": "ipython3",
"version": "3.9.7"
}
},
"nbformat": 4,
"nbformat_minor": 5
}

```

Statistics Assessment 3

1. Which of the following is the correct formula for total variation?

Ans. (b) Total Variation = Residual Variation + Regression Variation

2. Collection of exchangeable binary outcomes for the same covariate data are called outcomes.

Ans. © binomial

3. How many outcomes are possible with Bernoulli trial?

Ans. (a) 2

4. If H_0 is true and we reject it is called

Ans. (a) Type-I error

5. Level of significance is also called:

Ans. (a) Power of the test - alpha

6. The chance of rejecting a true hypothesis decreases when sample size is:

Ans. (a) Decreases

7. Which of the following testing is concerned with making decisions using data?

Ans. (b) Hypotheses

8. What is the purpose of multiple testing in statistical inference?

Ans. © Minimize false negatives

9. Normalized data are centred at and have units equal to standard deviations of the original data:

Ans. (a) 0

10. What Is Bayes' Theorem?

Ans. The Bayes theorem in Probability determines a given event's conditional probability using a mathematical formula.

11. What is z-score?

Ans. In statistics, Z-scores are used to calculate an observation's divergence from the mean value of the group. Z-scores tell statisticians and traders if a certain score is normal for a given data set or unusual.

12. What is t-test?

Ans. A t-test is a statistical test used to compare two groups' means. It is frequently used in hypothesis testing to assess whether a procedure or treatment has an effect on the population of interest, or whether two groups vary.

13. What is percentile?

Ans. A percentile is a concept in statistics that defines how a score relates to other scores from the same collection. While there is no uniform definition of percentile, it is often stated as the proportion of values in a set of data scores that are less than a certain value.

14. What is ANOVA?

Ans. An ANOVA test is a statistical test that is used to evaluate if there is a statistically significant difference between two or more category groups by testing for mean differences using variance. Another important aspect of ANOVA is that it divides the independent variable into two or more groups.

15. How can ANOVA help?

ANOVA is useful for comparing three or more variables. It is analogous to performing repeated two-sample t-tests. However, it produces fewer type I errors and is applicable to a wide range of problems. ANOVA groups differences by comparing the means of each group and involves dispersing variance across several sources.