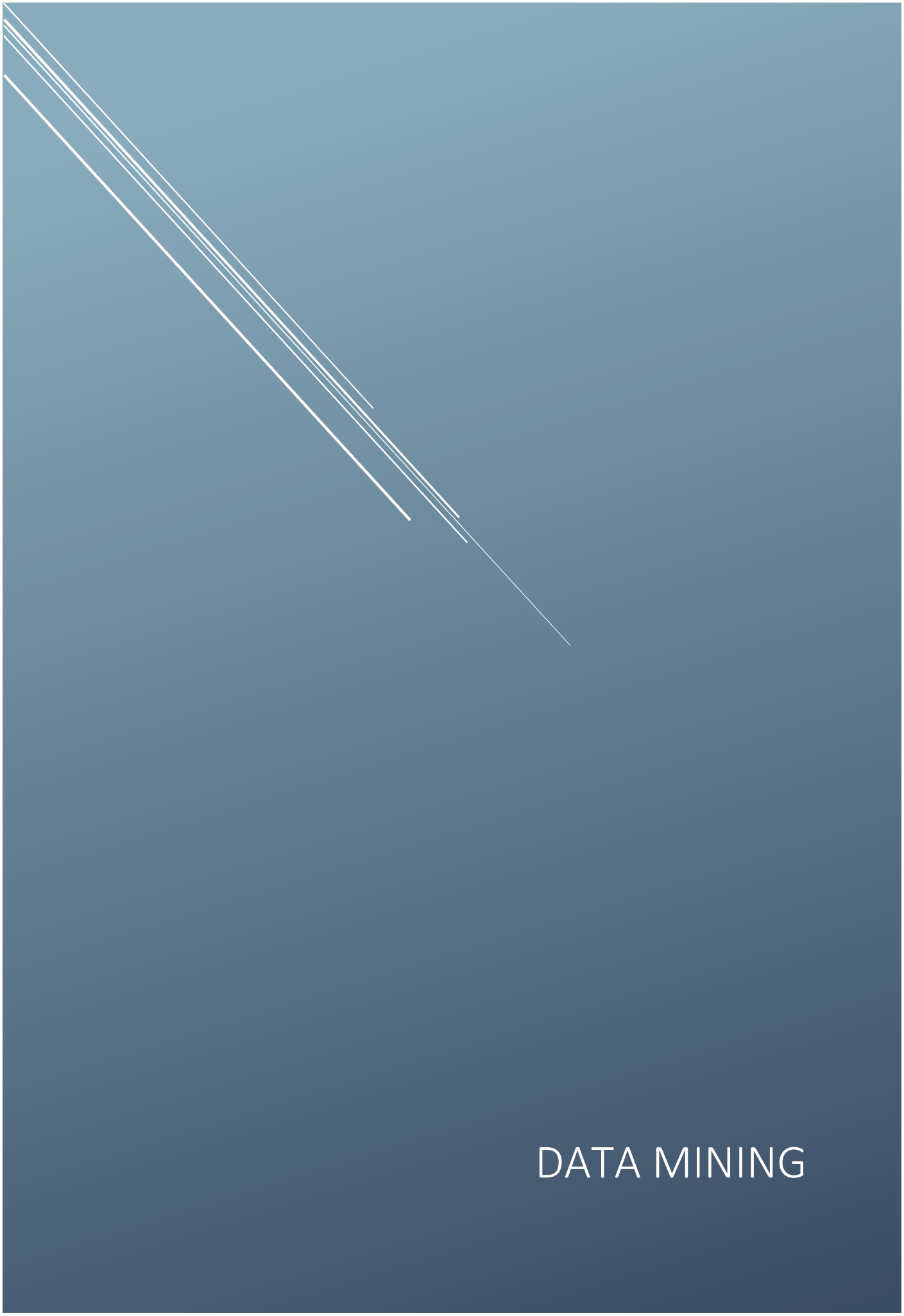
DATA MINING

Data Mining Assignments

اسم الطالب /

عبدالرحمن عبدالجليل عبدالرحمن

اشراف الدكتورة /

هبة المروعي

Spearman’s Rank Correlation

Introduction:

Before learning about Spearman’s correllation it is important to understand Pearson’s correlation which is a statistical measure of the strength of a linear relationship between paired data. Its calculation and subsequent significance testing of it requires the following data assumptions to hold:

- interval or ratio level;
- linearly related;
- bivariate normally distributed.

If your data does not meet the above assumptions then use Spearman’s rank correlation!

Monotonic function:

To understand Spearman’s correlation it is necessary to know what a monotonic function is. A monotonic function is one that either never increases or never decreases as its independent variable increases. The following graphs illustrate monotonic functions:

What Is Monotonic Function?

To understand Spearman's rank correlation, it is important to understand monotonic function. A monotonic function is one that either never increases or never decreases as its independent variable changes.

The following graph illustrates the monotonic function:

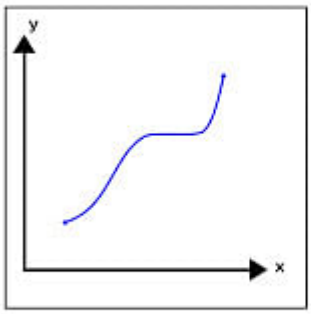

Figure 1 - A Monotonically Increasing function

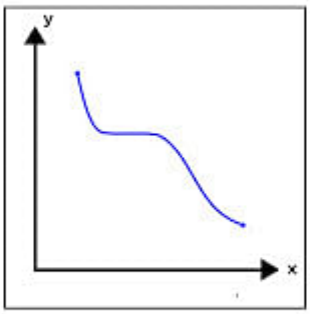

Figure 2 - A Monotonically decreasing function

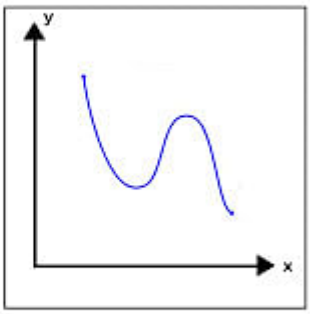

Figure 3 - A function that is not Monotonic

- Monotonically Increasing: As the variable X increases, the variable Y never decreases.
- Monotonically Decreasing: As the variable X increases, the variable Y never increases.
- Not Monotonic: As the X variable increases, the Y variable sometimes decreases and sometimes increases.

Spearman’s Rank Correlation

Spearman’s rank correlation measures the strength and direction of association between two ranked variables. It basically gives the measure of monotonicity of the relation between two variables i.e. how well the relationship between two variables could be represented using a monotonic function.

The formula for Spearman’s rank coefficient is:

$$\rho = 1 - \frac{6\Sigma\, d_i^2}{n(n^2 - 1)}$$

ρ = Spearman’s rank correlation coefficient

di = Difference between the two ranks of each observation

n = Number of observations

The Spearman Rank Correlation can take a value from +1 to -1 where,

- A value of +1 means a perfect association of rank
- A value of 0 means that there is no association between ranks
- A value of -1 means a perfect negative association of rank

Let’s understand the concept better with the help of an example.

Example of Spearman's Rank Correlation

Consider the score of 5 students in Maths and Science that are mentioned in the table.

Students	Maths	Science
A	35	24
B	20	35
C	49	39
D	44	48
E	30	45

Step 1: Create a table .

Step 2: Rank both the data in descending order. The highest marks will get a rank of 1 and the lowest marks will get a rank of 5.

Step 3: Calculate the difference between the ranks (d) and the square value of d.

Step 4: Add all your d square values.

Students	Maths	Rank	Science	Rank	d	d square
A	35	3	24	5	2	4
B	20	5	35	4	1	1
C	49	1	39	3	2	4
D	44	2	48	1	1	1
E	30	4	45	2	2	4
						14

Step 5: Insert these values into the formula.

$$\rho = 1 - \frac{6\Sigma\, d_i^2}{n(n^2 - 1)}$$

= 1 - (6 * 14) / 5(25 - 1)

= 0.3

The Spearman's Rank Correlation for the given data is 0.3. The value is near 0, which means that there is a weak correlation between the two ranks.