**Guía de ejercicios Capítulo IV**

**Reglas de Asociación**

1. **Objetivos de la Unidad**

| **#** | **Descripción** | **Preguntas** |
| --- | --- | --- |
| 1. | Comprender el entorno de problemas y la génesis de las reglas de asociación. |  |
| 2. | Conocer las definiciones formales de las reglas de asociación. |  |
| 3. | Comprender y manipular los parámetros necesarios para evaluar en forma cualitativa una regla. |  |
| 4. | Comprender los problemas combinatorios subyacentes a la búsqueda de relaciones frecuentes. |  |
| 5. | Evaluar las diferentes medidas de calidad de una regla de asociación. |  |
| 6. | Dominar los algoritmos para la búsqueda de reglas frecuentes. |  |

1. **Preguntas**

P1.

Exprese la siguiente medida de calidad en la forma probabilística más simple e interpretarla ¿a qué medida conocida se parece?

P2.

Una parte de la medida Gini para evaluar reglas de asociación está definida como:

Determine su monotonicidad (monótona o anti-monótona) para el caso del soporte (con confianza constante) y para el caso de confianza (con soporte constante).

P3.

Un médico especialista en enfermedades broncopulmonares infantiles tiene la convicción de que la estacionalidad del clima en Santiago influye fuertemente en el agravamiento de los infantes alérgicos. Para evaluar su hipótesis ha recolectado datos de los infantes con síntomas alérgicos e infantes normales y ha constatado sus ingresos a salas IRA (infección respiratoria aguda) de la ciudad de Santiago. Muestre una base de datos codificada para la aplicación del método de reglas de asociación, que tenga la particularidad de que cuando se agrega la variable estacionalidad (importa la instancia invierno) la confianza aumenta.

P4.

Un analista de marketing tiene la hipótesis que existe una fuerte relación (más del 50%) entre personas que están suscritas al diario *El Mercurio* y *Estrategia*, con las personas que toman sus vacaciones en el extranjero y en el sur del país. Para comprobar esta hipótesis posee las bases de datos de las suscripciones a los periódicos y de la empresa *Latam*.

Construya una base de datos de no más de 10 ejemplos que compruebe la hipótesis. Considerando un soporte de más del 10%, cuando el universo está constituido por las personas comunes a las tres bases de datos.

1. **Respuestas**

P2.

R =

Confianza constante: Considerando una confianza constante, el lado derecho de la ecuación siempre será positivo constante, mientras que el izquierdo, al tener el soporte de a, sería anti-monótona.

||

||

Soporte constante: Considerando un soporte constante, dado que cuando la confianza tome valores entre 0 y 1, se perdería la monoticidad (cuando la conf es 1, el lado derecho vale 0, cuando la conf es 0, el lado derecho es 1, cuando la confianza es 0.5, el lado derecho en 0.25)

||

||

||

**P3.**

R =

v1 = Tos v2 = Mucosidad v3 = Dificultad respiratoria v4 = Dolor en el pecho

v5 = Presión sanguínea anormal v6 = Estación invierno

A1 = [v1, v2, v3, v4, v5]

A2 = [v1, v2, v3, v4, v5, v6]

B = Alergia

**Tabla sin considerar la estación**

| **Paciente** | v1 | v2 | v3 | v4 | v5 | B |
| --- | --- | --- | --- | --- | --- | --- |
| **x1** | 1 | 1 | 1 | 1 | 1 | 1 |
| **x2** | 1 | 1 | 0 | 1 | 1 | 0 |
| **x3** | 0 | 0 | 0 | 0 | 0 | 0 |
| **x4** | 1 | 1 | 1 | 1 | 1 | 0 |
| **x5** | 0 | 1 | 1 | 1 | 0 | 1 |

Soporte(A1, B) = 1

Soporte(A1) = 2

Confianza = Soporte(A1, B) / Soporte(A1)

= 1 / 2 = 0.5

**Tabla considerando la estación**

| Paciente | v1 | v2 | v3 | v4 | v5 | v6 | B |
| --- | --- | --- | --- | --- | --- | --- | --- |
| x1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| x2 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| x3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| x4 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| x5 | 0 | 1 | 1 | 1 | 0 | 1 | 1 |

Soporte(A2, B) = 1

Soporte(A2) = 1

Confianza(A2, B) = Soporte(A2, B) / Soporte(A2)

= 1 / 1 = 1

Dado que en la tabla 2 se agregó la variable estación, y que esta hizo disminuir el Soporte(A2) (dado que para el paciente 4 ya no se cumplían todos los v), es que se puede decir que:

Confianza(A2,B) > Confianza(A1, B)