



INSTITUTO TECNOLÓGICO BELTRÁN
Centro de Tecnología e Innovación

Ejercicios

Trabajo Práctico N.º 2

Algoritmo de FOIL



Nombre: Coral Tolazzi
Tema: Algoritmo de FOIL
Profesora: Yanina Ximena Scudero
Cuatrimestre y Año: 2 Cuatrimestre del 2025

Instituto tecnológico Beltrán
Procesamiento de Aprendizaje Automático

Ejercicio 1: Identificar empleados en formación

Este ejercicio permite practicar cómo inducir reglas lógicas simples a partir de diferencias entre ejemplos positivos y negativos, tal como lo hace el algoritmo FOIL en su forma más básica. Se tiene un conjunto de datos con personas que trabajan en una empresa. Algunas están en formación (aprendices) y otras no. Cada persona tiene atributos como:

edad

departamento

nivel_educativo

en_formacion (booleano: True o False)

Dataset de ejemplo:

```
datos = [{"edad": 22, "departamento": "IT", "nivel_educativo": "terciario", "en_formacion": True},
{"edad": 24, "departamento": "IT", "nivel_educativo": "universitario", "en_formacion": True},
{"edad": 21, "departamento": "RRHH", "nivel_educativo": "terciario", "en_formacion": True},
{"edad": 35, "departamento": "IT", "nivel_educativo": "universitario", "en_formacion": False},
{"edad": 40, "departamento": "Finanzas", "nivel_educativo": "maestría", "en_formacion": False},
{"edad": 29, "departamento": "RRHH", "nivel_educativo": "universitario", "en_formacion": False},
{"edad": 23, "departamento": "IT", "nivel_educativo": "terciario", "en_formacion": True},
{"edad": 38, "departamento": "Finanzas", "nivel_educativo": "universitario", "en_formacion": False}]
```

Objetivo del ejercicio:

1. Separar los ejemplos positivos (`en_formacion == True`) y negativos (`en_formacion == False`).
2. Para cada atributo, identificar los valores que aparecen en los positivos pero no en los negativos.
3. Desarrollar un programa en Python que muestre la Regla inducida para identificar a personas en formación:

Regla inducida para identificar a personas en formación:

```
Regla inducida para identificar a un alumno:
- edad debe ser uno de: [24, 21, 22, 23]
- nivel educativo debe ser uno de: ['terciario']
```

Respuestas a las preguntas guía:

¿Qué valores de departamento aparecen solo en los positivos?

No hay ningún valor en departamento que solo aparece en los positivos.

¿Qué niveles educativos son comunes en los positivos pero no en los negativos?

el unico nivel educativo que es comun en los positivos pero no en los negativos es el terciario.

¿Qué edades aparecen solo en los positivos?

Las edades que solo aparecen en las positivas son 21, 22, 23 y 24.

Ejercicio 2:

a) Desarrollar un programa en Python que obtenga el cálculo del FOIL Gain para la condición `nivel_educativo == 'terciario'`, con la siguiente salida:

P (positivos antes) = 4

N (negativos antes) = 4

p (positivos después) = 3

n (negativos después) = 0
 $p / (p + n) = 1.000$ $P / (P + N) = 0.500$
 $\log_2(p / (p + n)) = 0.000$
 $\log_2(P / (P + N)) = -1.000$
 FOIL Gain = 3.000 b)

Realizar el cálculo de FOIL Gain manual para comprobar los cálculos. Condición: nivel_educativo == 'terciario' Valores utilizados:

Término	Valor
P (positivos antes)	4
N (negativos antes)	4
p (positivos después)	3
n (negativos después)	0

1. Fracción después de aplicar la condición:

$$p/(p+n) = 1.000$$

2. Fracción antes de aplicar la condición:

$$P/(P+N) = 4 / (4+4) = 4/8 = 0.500$$

3. Logaritmos:

$$\log_2(p/(p+n)) = \log_2(3/(3+0)) = \log_2(1.000) = 0.000$$

$$\log_2(P/(P+N)) = \log_2(4/(4+4)) = \log_2(0.500) = -0.500$$

4. FOIL Gain: 3.000

5. Interpretación:

Esta condición cubre la mayoría de los positivos pero incluye un negativo, lo que la hace menos precisa que condiciones más estrictas (por ejemplo, nivel educativo = terciario). Sin embargo, sigue teniendo buena ganancia de información, por lo que FOIL podría considerarla si busca una regla más general.

c) Desarrollar un programa en Python que obtenga el cálculo del FOIL Gain para la condición edad ≤ 23 , con la siguiente salida:

P (positivos antes) = 4
 N (negativos antes) = 4
 p (positivos después) = 3
 n (negativos después) = 0
 $p / (p + n) = 1.000$
 $P / (P + N) = 0.500$
 $\log_2(p / (p + n)) = 0.000$
 $\log_2(P / (P + N)) = -1.000$
 FOIL Gain = 3 x (0.000 - (-1.000)) = 3 x 1 = 3.000

d)

Realizar el cálculo de FOIL Gain manual para comprobar los cálculos.

Condición: **edad \leq 23**

Término	Valor
P (positivos antes)	4
N (negativos antes)	4
p (positivos después)	3
n (negativos después)	0

1. Fracción después de aplicar la condición:

$$p/(p+n) = 1.000$$

2. Fracción antes de aplicar la condición:

$$P/(P+N) = 4 / (4+4) = 4/8 = 0.500$$

3. Logaritmos:

$$\log_2(p/(p+n)) = \log_2(3/(3+0)) = \log_2(1.000) = 0.000$$

$$\log_2(P/(P+N)) = \log_2(4/(4+4)) = \log_2(0.500) = 0.500$$

4. FOIL Gain: 3.000

5. Interpretación:

Esta condición tiene los mismo resultados que la anterior condición