Tarea 09

Nombres y código:

Ruiz Ortiz Cesar Adrián 208020817

Padilla Martin Juan Pablo 217294261

Álvarez Gutiérrez David Alejandro 217294016

Sección: D01

Materia: MINERIA DE DATOS

Profesor: ISRAEL ROMAN GODINEZ

CARRERA: INGENIERIA EN INFORMATICA

CICLO ESCOLAR: 2020 A



Descripción del problema

- 1. Para la Tabla 1, construya un nuevo conjunto de datos que correspondan con las siguientes transformaciones.
- a. Normalice el atributo Edad para que esté en el rango de [-1, 1] considerando que la edad está acotada en el rango de edades de 1 a 110 años.

RESPUESTA:

Técnica: normalización min-max

Formula:
$$v'_i = \frac{v_i - min_A}{max_A - min_A} (new_max_A - new_min_A) + new_min_A.$$

			х	У			Z	z(nmax-nmin)+nmin
vi	min	max	vi-minA	maxA-minA	nuevo min	nuevo max	x/y	v´í
37	1	110	36	109	-1	1	0.330275	-0.339449541
41	1	110	40	109	-1	1	0.366972	-0.266055046
56	1	110	55	109	-1	1	0.504587	0.009174312
57	1	110	56	109	-1	1	0.513761	0.027522936
62	1	110	61	109	-1	1	0.559633	0.119266055
63	1	110	62	109	-1	1	0.568807	0.137614679
67	1	110	66	109	-1	1	0.605505	0.211009174
67	1	110	66	109	-1	1	0.605505	0.211009174

b. Normalice el atributo de Colesterol considerando que no se conocen los valores mínimos ni máximos que se pueden alcanzar con esta medición.

RESPUESTA:

Técnica: normalización Z-score

Formula:
$$v_i' = \frac{v_i - \bar{A}}{\sigma_A}$$
,

			×	x/desv std
vi	media	desv std	vi-media	ví
204	254.1834	46.33419	-50.1834143	-1.0830753
229	254.1834	46.33419	-25.1834143	-0.5435169
233	254.1834	46.33419	-21.1834143	-0.4571875
236	254.1834	46.33419	-18.1834143	-0.3924405
250	254.1834	46.33419	-4.18341425	-0.0902878
268	254.1834	46.33419	13.8165857	0.29819418
286	254.1834	46.33419	31.8165857	0.68667621
354	254.1834	46.33419	99.8165857	2.15427498

c. Convierta los atributos Fumador, Escolaridad e Ingreso en su correspondiente representación numérica.

RESPUESTA:

Fumador

Técnica: binario a numérico

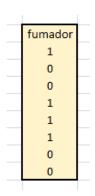
• Atributo binario:

Fumador = {si,no}

Transformación

Fumador= $si \rightarrow 1$

Fumador= no ->= 0



Escolaridad

Técnica: ordinal a numérico

• dominio:

Escolaridad = {secundaria, preparatoria, universidad, maestria}

• asignación de valor consecutivo

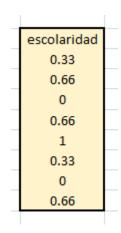
{ secundaria -> 1, preparatoria->2, universidad->3, maestria->4 } \rightarrow z = {1,2,3,4}

normalización:

formula: $Z_i' = \frac{z_i - 1}{|Z| - 1}$

1	-1	3	0	0	
2	-1	3	1	0.333333	
3	-1	3	2	0.666667	
4	-1	3	3	1	

 $Z'=\{0,0.33,0.66,1\}$



Ingreso

Técnica: ordinal a numérico

• dominio:

Escolaridad = $\{x \le 50k, 50k < X \le 100k, X > 100k \}$

• asignación de valor consecutivo

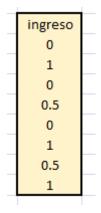
{ $x \le 50k -> 1, 50k < X \le 100k -> 2, X > 100k -> 3$ } \rightarrow $z = \{1,2,3,\}$

formula: $Z_i' = \frac{z_i - 1}{|Z| - 1}$ }

normalización

1	-1	2	0	0	
2	-1	2	1	0.5	
3	-1	2	2	1	

$Z'=\{0,0.5,1\}$



Edad	Fumador	Colesterol	Escolaridad	Ingreso
63	Si	233	Preparatoria	X <=50K
67	No	286	Universidad	X > 100K
67	No	229	Secundaria	X <=50K
37	Si	250	Universidad	50K < X <= 100K
41	Si	204	Maestría	X <=50K
56	Si	236	Preparatoria	X > 100k
62	No	268	Secundaria	50K < X <= 100K
57	No	354	Universidad	X > 100K

Tabla 1