

Inteligencia Artificial

Instituto Nacional de Astrofísica, Óptica y Electrónica

Presents:

MSc. Mireya Lucia Hernández Jaimes

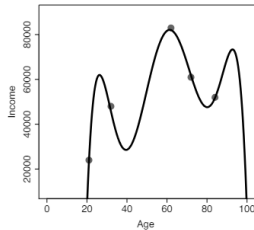
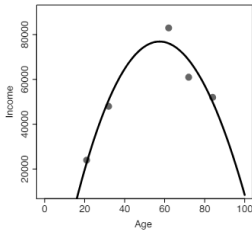
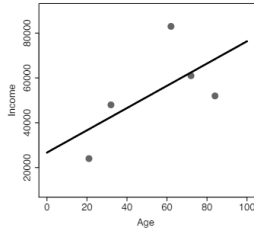
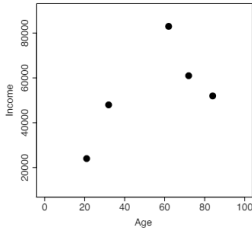
① Ejercicios

Indique si cada característica descriptiva contiene datos numéricos, ordinales, nominales, binarios o textuales.

ID	OCCUPATION	GENDER	AGE	MOTOR VALUE	POLICY TYPE	PREF CHANNEL
1	lab tech	female	43	42,632	planC	sms
2	farmhand	female	57	22,096	planA	phone
3	biophysicist	male	21	27,221	planA	phone
4	sheriff	female	47	21,460	planB	phone
5	painter	male	55	13,976	planC	phone
6	manager	male	19	4,866	planA	email
7	geologist	male	51	12,759	planC	phone
8	messenger	male	49	15,672	planB	phone
9	nurse	female	18	16,399	planC	sms
10	fire inspector	male	47	14,767	planC	email

Regresión lineal

Todos los modelos son algo consistentes con el datos de entrenamiento, pero ¿cuál es mejor?



You are given a vector a measurements x and true values y

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad y = \begin{bmatrix} 1.5 \\ 2 \\ 2.5 \end{bmatrix}$$

- a) Plot y and x as points.
- b) If we start with $\theta^0=0$ and $\theta^1=0$, what is the initial value for the loss function?
- c) Compute the next estimate of θ^0 and θ^1 , after 1 iteration of gradient descent.

Se ha construido un modelo de regresión lineal multivariante para predecir la carga de calefacción en un edificio residencial basándose en un conjunto de características descriptivas que describen las características del edificio.

$$\begin{aligned}\text{HEATING LOAD} = & -26.030 + 0.0497 \times \text{SURFACE AREA} \\ & + 4.942 \times \text{HEIGHT} - 0.090 \times \text{ROOF AREA} \\ & + 20.523 \times \text{GLAZING AREA}\end{aligned}$$

Utilice este modelo para hacer predicciones para cada una de las instancias de consulta que se muestran en la siguiente tabla.

ID	SURFACE AREA	HEIGHT	ROOF AREA	GLAZING AREA
1	784.0	3.5	220.5	0.25
2	710.5	3.0	210.5	0.10
3	563.5	7.0	122.5	0.40
4	637.0	6.0	147.0	0.60

Se ha construido un modelo de regresión logística multivariante para predecir la propensión de los compradores a realizar una compra repetida de un obsequio que reciben. El departamento de marketing está utilizando este modelo para determinar a quién se le debe dar el obsequio. Los pesos en el modelo entrenado se muestran en la siguiente tabla.

Feature	Weight
Intercept ($w [0]$)	-3.82398
AGE	-0.02990
SOCIO ECONOMIC BAND B	-0.09089
SOCIO ECONOMIC BAND C	-0.19558
SHOP VALUE	0.02999
SHOP FREQUENCY	0.74572

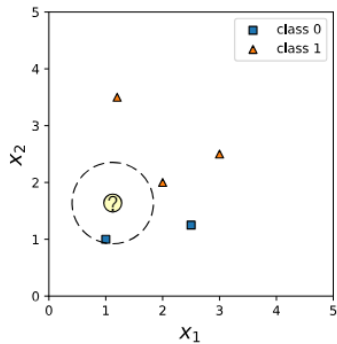
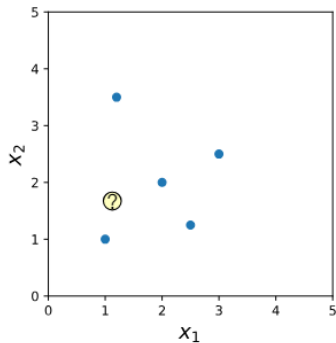
Utilice este modelo para hacer predicciones para cada una de las siguientes instancias de consulta.

ID	AGE	SOCIO ECONOMIC	SHOP	SHOP
		BAND	FREQUENCY	VALUE
1	56	b	1.60	109.32
2	21	c	4.92	11.28
3	48	b	1.21	161.19
4	37	c	0.72	170.65

Árboles de decisión utilizando entropía e información de ganancia

Construye un árbol de decisión ID3 para el siguiente dataset:

Day	Outlook	Temp.	Humidity	Wind	Decision
1	Sunny	85	85	Weak	No
2	Sunny	80	90	Strong	No
3	Overcast	83	78	Weak	Yes
4	Rain	70	96	Weak	Yes
5	Rain	68	80	Weak	Yes
6	Rain	65	70	Strong	No
7	Overcast	64	65	Strong	Yes
8	Sunny	72	95	Weak	No
9	Sunny	69	70	Weak	Yes
10	Rain	75	80	Weak	Yes
11	Sunny	75	70	Strong	Yes
12	Overcast	72	90	Strong	Yes
13	Overcast	81	75	Weak	Yes
14	Rain	71	80	Strong	No



La siguiente tabla enumera un conjunto de datos que se utilizó para crear un modelo de vecino más cercano que predice si será un buen día para surfear.

ID	WAVE SIZE (FT)	WAVE PERIOD (SECS)	WIND SPEED (MPH)	GOOD SURF
1	6	15	5	yes
2	1	6	9	no
3	7	10	4	yes
4	7	12	3	yes
5	2	2	10	no
6	10	2	20	no

Suponiendo que el modelo utiliza la distancia euclidiana para encontrar el vecino más cercano ($K=1$), ¿qué predicción devolverá el modelo para cada una de las siguientes instancias de consulta?

ID	WAVE SIZE (FT)	WAVE PERIOD (SECS)	WIND SPEED (MPH)	GOOD SURF
Q1	8	15	2	?
Q2	8	2	18	?
Q3	6	11	4	?

La característica objetivo es el Índice de Percepción de la Corrupción (CP1). El CPI mide los niveles percibidos de corrupción en el sector público de los países y oscila entre 0 (altamente corrupto) y 100 (muy limpio)

COUNTRY ID	LIFE EXP.	TOP-10 INCOME	INFANT MORT.	MIL. SPEND	SCHOOL YEARS	CPI
Afghanistan	59.61	23.21	74.30	4.44	0.40	1.5171
Haiti	45.00	47.67	73.10	0.09	3.40	1.7999
Nigeria	51.30	38.23	82.60	1.07	4.10	2.4493
Egypt	70.48	26.58	19.60	1.86	5.30	2.8622
Argentina	75.77	32.30	13.30	0.76	10.10	2.9961
China	74.87	29.98	13.70	1.95	6.40	3.6356
Brazil	73.12	42.93	14.50	1.43	7.20	3.7741
Israel	81.30	28.80	3.60	6.77	12.50	5.8069
U.S.A	78.51	29.85	6.30	4.72	13.70	7.1357
Ireland	80.15	27.23	3.50	0.60	11.50	7.5360
U.K.	80.09	28.49	4.40	2.59	13.00	7.7751
Germany	80.24	22.07	3.50	1.31	12.00	8.0461
Canada	80.99	24.79	4.90	1.42	14.20	8.6725
Australia	82.09	25.40	4.20	1.86	11.50	8.8442
Sweden	81.43	22.18	2.40	1.27	12.80	9.2985
New Zealand	80.67	27.81	4.90	1.13	12.30	9.4627

Usaremos Rusia como nuestro país de consulta para esta pregunta. La siguiente tabla enumera las características descriptivas de Rusia.

COUNTRY ID	LIFE EXP.	TOP-10 INCOME	INFANT MORT.	MIL. SPEND	SCHOOL YEARS	CPI
Russia	67.62	31.68	10.00	3.87	12.90	?

¿Qué valor arrojaría un modelo de predicción de los tres vecinos más cercanos utilizando la distancia euclidiana para el CPI de Rusia?