

[Atrás](#)

Puede previsualizar este cuestionario, pero si éste fuera un intento real, podría ser bloqueado debido a:

Este cuestionario no está disponible en este momento

Pregunta 1

Sin responder aún Puntúa como 6,0

Selecciones todas las opciones verdaderas, una respuesta Falsa seleccionada genera menos 0.5 puntos.

- a. Una matriz estandarizada tiene una media igual a cero y la desviación estándar es uno. *Si $(X-\bar{X})/\sigma \approx N(0,1)$*
- b. Una matriz centrada es la que la media es cero y la desviación es 1, *NO solo que lo constante es 0*
- c. Un histograma de datos con outliers debe tener un menor numero de bins para ver la verdadera distribución de los datos. *No, al contrario se debe tener mas grupos para identificar*
- d. Una correlación negativa mayor a 0.4 entre X y Y, significa que un cambio en la variable X genera una variación de al menos 40% de la variable Y. *(40% de la varianza de Y esta explicada por X)*
- e. Cuando la mediana es menor que la media se espera que el histograma tenga mas observaciones al lado izquierdo de la media. *Si, la media esta afectada por el outlier hacia la derecha*
- f. Si la mediana esta a cuatro desviaciones estándares de la media es porque existe al menos un outlier *Si, los outliers son datos > 30*
- g. Una variable con distribución normal se espera tenga la mediana igual a media, el valor del primer quartile esta muy cerca a la media menos 0.6 desviaciones estándar y el tercer quartile esta cerca a la media mas 0.6 desviaciones estándar *Si, es muy cerca a una distribución normal*
- h. La correlación entre la variable X y Y es diferente a la correlación de Y y X *NO, es una medida simétrica*

Pregunta 2

Sin responder aún Puntúa como 1,0

Cual es el numerador de la correlación de la siguiente tabla

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

x	3	7	11	15	19	23
y	-11	41	13	65	37	89

Respuesta:

$$n=6 \quad \sum x = 78 \quad \sum y = 239 \quad \sum xy = -33 + 287 + 193 + 975 + 703 + 2047 = 4122$$

$$r = \frac{6 \times 4122 - (78 \times 239)}{6 \times 1294 - 6084} = \frac{6480}{8006.79}$$

Pregunta 3

Sin responder aún Puntúa como 1,0

Cual es el resultado de la correlación de la siguiente tabla, y describa el significado del valor entre la variable x y y.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

x	3	7	11	15	19	23
y	-11	41	13	65	37	89

$$r = \frac{6480}{\sqrt{[6 \times 1294 - 6084][6 \times 15486 - 59756]}} = \frac{6480}{8006.79}$$

$$r = 0.8093$$

Pregunta 4

Sin responder aún Puntúa como 16,0

$$\sum_{i=0}^n \min_{\mu_j \in C} (\|x_i - \mu_j\|^2)$$

x1	x2	cluster	cluster segunda	última iteración
28	4	1	2	2
7	10	1	2	1
3	29	1	2	1
26	21	1	1	2
27	21	1	1	2
7	6	2	2	1
22	21	2	1	2
29	10	2	1	2
21	11	2	2	2
1	26	2	2	1
26	25	2	1	2

- a) Qual es el centroide de la iteración actual para el cluster 1 y 2? $\mu_1 = (18.2, 17)$ $\mu_2 = (17.66, 16.5)$
- b) Qual es el nuevo label para los puntos en la iteración 2? *ver tabla*
- c) Qual es el centroide para la iteración 2? $\mu_1 = (26, 19.6)$ $\mu_2 = (11.16, 14.33)$
- d) Cuantas iteraciones son necesarias para que el k-mean termine? 4
- e) Qual es el cluster final para cada punto? *ver tabla*

Pregunta 5

Sin responder aún Puntúa como 20,0

Utiliza los datos de las siguientes tablas para calcular PCA y reducir los valores a 1 Dimension.

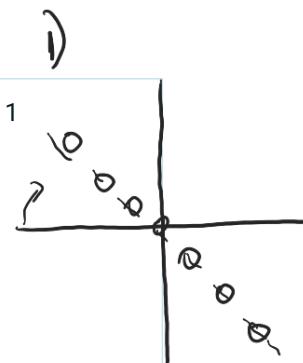
x_1	$\bar{x}_1 = 14, \sigma_1 = 2.73$	x_2	$\bar{x}_2 = 14, \sigma_2 = 2.73$
13		-13	
18		-18	
16		-16	
17		-17	
12		-12	
10		-10	
11		-11	
15		-15	

a) Cuál es la media de cada variable? y la desviación?

b) Cuáles son los valores de los eigen values?, y cuanta varianza es explicada por la primera componente?

c) Cuál es el valor del primer eigen vector?

d) Cuál es la proyección de estos valores en 1 dimensión.



$$\text{b)} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$\det \begin{pmatrix} 1-\lambda & -1 \\ -1 & 1-\lambda \end{pmatrix}$$

$$(1-\lambda)^2 - 1 = 0$$

$$1 - 2\lambda + \lambda^2 - 1 = 0$$

$$\lambda(\lambda - 2) = 0$$

$$\lambda_1 = 0, \lambda_2 = 2$$

$$\text{Para } \lambda_1 = 2$$

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$x_1 - x_2 = 2x_1 \Rightarrow x_1 = -x_2$$

$$-x_1 + x_2 = 2x_2 \Rightarrow$$

$$v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \|v_1\| = 1$$

$$\sqrt{x_1^2 + x_2^2} = 1 \Rightarrow 2x_1^2 = 1, x_1 = \sqrt{0.5}$$

$$v_1 = \begin{bmatrix} \sqrt{0.5} \\ -\sqrt{0.5} \end{bmatrix}$$

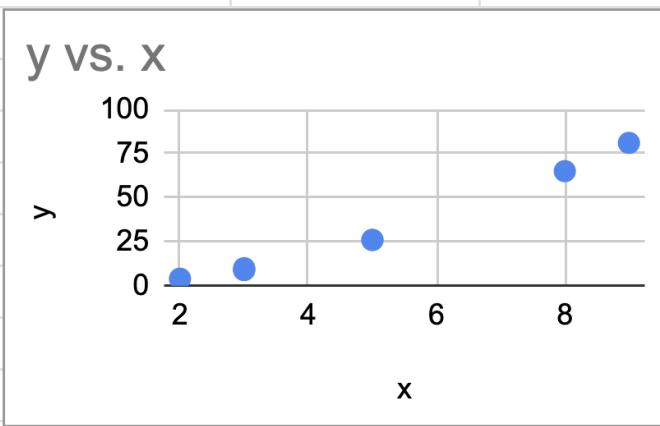
$$z = \begin{bmatrix} -0.36 & 0.36 \\ 1.46 & -1.46 \\ 0.73 & -0.73 \\ 1.09 & -1.09 \\ -0.73 & 0.73 \\ -1.46 & 1.46 \\ -1.09 & 1.09 \\ 0.36 & -0.36 \end{bmatrix} \begin{bmatrix} \sqrt{0.5} \\ -\sqrt{0.5} \end{bmatrix} = \begin{bmatrix} -0.5 \\ 2.0 \\ 1.0 \\ 1.54 \\ -1.03 \\ -2.06 \\ -1.54 \\ 0.51 \end{bmatrix}$$

Pregunta 6

Sin responder aún Puntúa como 20,0

Utilizando los datos de la siguiente tabla. Conteste las siguientes preguntas.

x	y	split
5	26	train
5	26	train
3	9	train
8	65	train
3	9	test
3	10	test
9	81	test
2	4	test
9	82	test
4	17	train
9	81	test
6	37	test
2	4	train
0	0	train



$$\bar{x} = 3.85 \quad \bar{y} = 21$$

$$\beta_1 = \frac{(5-3.8)(26-21)+\dots+(2-3.8)(0-21)}{38.8} = \frac{316}{38.8} = 8.13, \quad \beta_0 = 21 - 3.85 \cdot 8.13 = -10.36$$

- a) Calcular b_1 y b_0 , para crear la ecuación de la regresión (Solo utilizar train).
- b) Cual es el MSE y MAPE para train y test. $MSE_{train} = 43.73$ $MSE_{test} = 153.66$
- c) Cual es el R2, el RSS y el TSS. $MAPE_{train} = 0.30$ cero , $MAPE_{test} = 0.30$
- d) Que transformación puede aplicar a X para reducir el error?, cual es el nuevo r2?

$$RSS = 306.17 \quad TSS = 2876 \quad R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{306.17}{2876} = 0.8935$$

A ✓ B I ≡ ≡ % ⚡ H-P

la relacion entre X y y es cuadatica entonces transformar X^2 debe ayudar a tener un mejor r^2

$$\text{el nuevo } r^2 = 0.99$$

Pregunta 7

Sin responder aún Puntúa como 12,0

Utiliza 3NN vecinos para predecir el valor de y para los datos de test (Los vecinos solo deben ser los datos de train).

x	y	split	
2,4	5	26	train
2,4	5	26	train
1,3	3	9	train
2,9	8	65	train
	3	9	test 1 → $9+17+4 = 30/3 = 10$
	3	10	test 1 → 10
	9	81	test 2 → $65+26+26/3 = 39$
	2	4	test 3 → 10
	9	82	test → 39
1,3	4	17	train
	9	81	test → 39
	6	37	test 4 → $(26+26+65)/3 = 39$
1,3	2	4	train
0	0	0	train

a) Cual es el MSE y el MAPE?

b) Q modelo funciona mejor, la regresión o el 3NN?

$MSE = 774$ $MAPE = 0.46$
 La regresión tiene un $MAPE = 0.3$

