

1. Una regresión lineal es:
 - a. Un algoritmo de aprendizaje supervisado inflexible y paramétrico
2. Un sistema DFS tiene ventajas y desventajas. Estas son:
 - a. Tolerancia al fallo
 - b. Baja Consistencia
 - c. Alta concurrencia
3. Las características de un DFS son:
 - a. Permite tolerancia al fallo
 - b. Permite alta concurrencia
 - c. A mayor replicación menor consistencia
 - d. Divide los archivos en partes homogéneas en distintas máquinas
4. Un datasmart es:
 - a. Un subconjunto de un almacén de datos orientado a una línea de negocio específica. Los data marts contienen depósitos de datos resumidos recopilados para su análisis en una sección o unidad específica.
5. El problema de la latencia
 - a. La latencia es la demora dentro de un sistema basada en demoras en la ejecución de una tarea. Cuanto mas cerca este la respuesta de un cliente en el momento de la decisión, mas importa la latencia.
6. Los datos no estructurados pueden provenir de:
 - a. Fotos y Videos
 - b. Texto
 - c. Relaciones en redes sociales
7. El big data stack se refiere a:
 - a. La arquitectura de la Plataforma y programabilidad del big data
8. Para ejecutar la acción de contar palabras con MapReduce de Hadoop:
 - a. Primero carga el archivo en el sistema usando HDFS y luego ejecuta la aplicación de recuento de palabras basada en MapReduce sobre él.
9. Un sistema que abre la puerta con huella digital es:
 - a. Cognitive system (sistema cognitivo)

cta

La **varianza** se refiere al error en el resultado debido al cambio del **conjunto de entrenamiento**.

El **sesgo** se refiere al error en el resultado debido al cambio en la **forma de la función**

Y en **métodos más flexibles**, la varianza aumentará y el sesgo disminuirá. por lo tanto, en muchas ocasiones, un método simple supera a los más flexibles

10.

☐ Primero clusterizar, Luego utilizar un algoritmo supervisado de predicción que determine con precisión las medidas de la prenda personalizada más adecuada y luego clasificar por KNN

recta

☐ Existen varias alternativas si bien podríamos primero clusterizar para encontrar grupos (si bien podría no ser necesario) y dentro de ellos utilizar KNN

☒ Existen varias alternativas si bien podríamos primero clusterizar para encontrar grupos (si bien podría no ser necesario) y dentro de ellos utilizar un algoritmo supervisado de clasificación.

11.

12. La escalabilidad horizontal:

- a. Implica agregar más máquinas, posiblemente menos potentes, a una red relativamente mas lenta. Muchos nodos de productos básicos en un clúster.

13. Serialization (serialización) es:

- a. La serialización es el proceso de traducir las estructuras de datos o el estado de los objetos en forma binaria o textual para transportar los datos a través de la red o almacenarlos en algún almacenamiento persistente

14. Virtualización es:

- a. Hace posible asignar tareas a un conjunto de recursos virtuales

15. Las inferencias están relacionadas con:
- ¿Cuál es la relación entre la respuesta y cada predictor?
 - ¿Qué predictores están asociados con la respuesta?
 - ¿Qué predictores no están asociados con la respuesta?
16. ¿Qué pasaría probablemente si el parámetro alfa fuera 0,1?
- Que el método no convergería
17. ¿Qué pasaría si con los parámetros de inferencia si cambio o añadido alguno de esos datos de aprendizaje?
- Que los parámetros de inferencia cambiarían, ajustándose mejor a la nueva muestra.
18. K-means y KNN, ¿Cuáles de las siguientes afirmaciones son Correctas referidas a las dos metodologías?
- KNN es supervisado y K-means no supervisado
 - El propósito de K-means es encontrar grupos y K-NN es clasificar un nuevo elemento dado
19. Las inferencias están relacionadas con
- Que predictores no están asociados con la respuesta
 - Que predictores están asociados con la respuesta
20. El auge de la IA ha tenido lugar porque
- Porque el rendimiento de las redes neuronales no aumenta cuando aumenta la cantidad de datos. Por el contrario, la IA tradicional no aumenta el rendimiento hasta cierto volumen de datos.
21. La Ley de Moore se refiere a la percepción de que la cantidad de transistores en un microchip se duplica cada ____ años, aunque el costo de las computadoras se reduce a la mitad
- 2 años
22. Data Munging es el proceso inicial de refinación de datos en bruto en contenido o formatos mas adecuados para el consumo por sistemas y usuarios de corriente. Este concepto incluye
- Enriquecimiento de datos
 - Validacion de datos
 - Transformación de datos e integración
23. La persistencia garantiza que los datos almacenados en una base de datos no serán ____ sin permisos y que estarán disponibles si son importantes para el negocio.
- Variados/ Modificados

24. A lo largo de los años el lenguaje de consulta estructurado (SQL) ha evolucionado al mismo ritmo que la tecnología RDBMS y es el mecanismo más utilizado para crear, consultar, mantener y operar bases de datos relacionales. Estas tareas se denominan CRUD:
- Update (actualizar)
 - Create (crear)
 - Retrieve (extraer)
25. Mientras que RDBMS usa ACID como mecanismo para garantizar la consistencia de los datos, DBMS no relacional usa ____
- Base
26. Datos en movimiento: analizados a medida que se generan. De lo contrario, los registros se recopilan en una ubicación de búfer (se crea lote de registros) y se envían a la base de datos en fragmentos discretos a intervalos periódicos. ____ recopilados antes del análisis
- Data at rest
27. Streaming computing está diseñado para manejar un flujo continuo de una gran cantidad de datos no estructurados. Por el contrario, ____ generalmente trata con algunas variables que deben correlacionarse con un proceso comercial específico.
- CEP
28. Un ____ de datos es un tipo de base de datos que integra copias de datos de transacciones de sistemas de origen dispares y los aprovisiona para uso analítico
- Datawarehouse
29. Los datalakes garantizan que todos los datos se almacenen para un uso potencialmente desconocido más adelante. Esquema en ____
- Read/ Lectura
30. Las tecnologías emergentes de memoria principal no volátil por lo general ofrecen una memoria de ____ mucho más alta, un costo por bit y un consumo de energía en espera mucho más bajos que la DRAM
- Densidad
31. En la arquitectura abajo indicada, el datawarehouse se usa porque:
- Para preparar data marts para realizar diferentes análisis para diferentes departamentos en una empresa
 - Tenemos varias bases de datos transaccionales y hay que combinar y preparar para el análisis

32. En la arquitectura de la figura se combinan
- a. Streaming data y data at rest
 - b. Datos estructurados y no estructurados
 - c. Bases de datos distribuidas sobre hadoop y datawarehouses
 - d. Datos internos y externos
33. Si estamos interesados al analizar principalmente en _____ entonces los modelos restrictivos son mucho más interpretables
- a. Inferencia
34. La actuación de la IA
- a. Aumenta mas con larga redes neuronales y grandes cantidades de datos
 - b. Aumenta cuanto la cantidad de data aumenta hasta cierto punto usando IA tradicional
 - c. Ha despegado últimamente debido a el aumento de redes neuronales y Deep learning
35. Valencia:
- a. Se refiere a la conectividad
 - b. Una mayor valencia da lugar a mayor enlace y produce mejor conectividad
36. A _____ es un almacén masivo de datos con gran poder de procesamiento y habilidad para manejar una gran cantidad de numero de concurrencias, gestión de datos y tareas analíticas. En un data lake los datos no se almacenan en un almacén a menos que estén en uso.
- a. Data lake
37. La precisión de Y como predicción de Y depende de dos magnitudes, que denominamos
- a. Error reducible y error no reducible
38. Un DFS consiste básicamente en
- a. Un commodity cluster
 - b. Una partición de un gran conjunto de datos y la replicación de cada una de esas partes en varias computadoras
39. Un banco que actualiza todas las transacciones todos los días por la noche realiza un
- a. Proceso Batch
40. La Ps del Big Data son
- a. Personas, Procesos, Propósito, Plataforma y Programabilidad

41. Las fuentes de datos son
- a. Personas, Máquinas y empresas (corporations)
42. Para usar MapReduce para contar palabras usando Cloudera
- a. Primero subimos el archivo en el sistema usando HDFS y después ejecutamos la aplicación base de contar palabras de MapReduce sobre el archivo
43. Análisis cognitivo se refiere a
- a. Un modelo de análisis de datos que responde y actúa automáticamente
44. Input-data (datos de entrada) se refiere a
- a. Cualquier dato que un humano pueda introducir en un ordenador
45. La capa 0 de la pila del big data se refiere a
- a. Infraestructura física—hardware, network
46. Si cambiamos el training set y los parámetros calculados para una predicción cambian mucho, tenemos
- a. Gran Variación (alta variación)
47. The reducible error in a prediction se refiere a/ depende de
- a. Depende de la precisión de la función
48. Varianza, sesgo
- a. La varianza se refiere al error en el resultado debido al cambio del conjunto de entrenamiento
Es sesgo se refiere al error en el resultado debido al cambio en la forma de la función
Y en métodos más flexibles, la varianza aumentará y el sesgo disminuirá, por lo tanto, en muchas ocasiones, un método simple supera a los más flexibles.
49. El Memorial de los Veteranos de Vietnam logra su fuerza visual y emocional mediante el micro/macro diseño
- a. Ordenando los nombres por la fecha de muerte en vez de alfabéticamente
50. Edward Tufte acuñó el término chartjunk para referirse a estas distracciones visuales. Se refiere a...
- a. Todos los elementos visuales de cuadros y gráficos que no sean necesarios para comprender la información representada en este gráfico o que distraigan al espectador de esta información

51. La estratificación y la separación suelen implicar la creación de una jerarquía de efectos visuales, posiblemente de información. ¿Cuántas capas pueden verse en las imágenes siguientes?
- a. Primer gráfico 2 capas: 2 triángulos Segundo gráfico 3 capas: 1 triángulo, 1 pseudo-triángulo irregular, círculos
52. Lecturas micro/macro...
- a. Los microdetalles permiten tanto macro
53. DDOS o ataque distribuido de denegación de servicio se refiere...
- a. Una saturación se consigue sobrecargando los servidores y normalmente el ataque proviene de varios puntos como la red de bots (máquinas)
54. ____ implica asegurarse de que el mensaje recibido no ha sido alterado o, al menos o al menos detectar que lo ha sido.
- a. Integridad de los mensajes
55. Sabemos que el mensaje "khuh zh duh" ha sido cifrado utilizando un desplazamiento de la Cifra César. Responde cuál es el número de rotaciones que descifran el alfabeto hacia la izquierda...
- a. (here we are)
56. El cifrado simétrico es...
- a. Una forma de criptosistema en la que el cifrado y el descifrado se realizan con la misma clave. También se conoce como cifrado convencional.
57. Usando el cifrado de Vigenère dar el texto cifrado de "SERENDIPITY" usando la clave LEMON...
- a. DIDS AOMBWGJ
58. Scytale es
- a. Cifrado monoalfabético por transposición
59. El principio de Kerckhoff postula que...
- a. un criptosistema debe seguir siendo seguro, aunque el adversario conozca cada detalle del sistema, excepto la clave
60. El único sistema de cifrado existente que, utilizado correctamente, es matemáticamente seguro, lo que significa que ninguna cantidad de tiempo o poder de procesamiento puede recuperar el texto plano del cifrado es...
- a. One-time pad (no se que de un solo uso)

61. La solución Kasiski para romper la clave Vigenère se basa en...
- a. Descubrir la longitud de la clave y, a continuación, descifrar varios
 - b. cifrados monoalfabéticos
62. Los tipos de intereses que podrían justificar el derecho a la intimidad son (seleccione todos los que sean se aplican) ...
- a. Power
 - b. Control
 - c. Derivate
63. La atomicidad se refiere a:
- a. Una transacción es "todo o nada" cuando es atómica. Si cualquier parte de la transacción o el sistema subyacente falla, toda la transacción falla.
64. Una API
- a. Puede ser horizontal o vertical
65. Si en un modelo lineal utilizamos un parámetro de aprendizaje muy grande para encontrar el error mínimo normalmente lo haríamos:
- a. no los encuentra porque el descenso de gradiente diverge
66. Si queremos comercializar diferentes bebidas en un mercado, para definir las mejores dos opciones posibles, lo haremos:
- a. ejecutar un K-means y desarrollar los centroides como representantes de los dos clusters
67. Seleccione los distintos niveles en los que la PNL realiza análisis sobre el texto
- a. Análisis sintáctico, análisis semántico, análisis a nivel del discurso, Análisis léxico/morfológico. (Las que pone "análisis")
68. Análisis del sentimiento
- a. se utiliza para identificar puntos de vista o emociones en el texto subyacente subyacentes. Algunas técnicas lo hacen clasificando el texto, por ejemplo, como subjetivo (opinión) u objetivo (hecho), utilizando técnicas de aprendizaje automático o PNL.
69. La distribución de la centralidad entre los vectores y la centralidad de los vectores propios:
- a. Ambos están muy sesgados
70. La visión dinámica de las redes incluye
- a. Evolución de redes (formación de redes), A través de la red (difusión).
71. Usando el Playfair Cypher con la frase HELLO WORLD -key: SPAIN
- a. LCKYGTVQTKAR

72. La corrección Wau de la correlación en la fórmula siguiente está considerando
- En qué medida relevante es la opinión de un usuario para otro.

Choropleths

Representar un mapa de regiones geográficas con una escala de colores para ver como varían los datos

PRACTICOS

What is Schelling's Segregation metric? 61.538%

In the figure below there are 24 blocks 10 people per block. Blue blocks are rich, yellow poor and green 50/50.



Given the ratings and correlations below calculate the rating predictions of PD3: 3.56

user-item	1.- Batman Begins	2.-Alice in wonderla nd	3.-Dumb and Dumber	4.- Equilibriu m	
User A	3	2	3	5	
User B	2	5	4	3	
User C	3	4	2		
User D	2	3		3	
User E	2	4	5		

user-user correlatio ns w	User A	User B	User C	User D	User E	
User A		-0.51	-0.87	0.19	-0.19	-1.38
User B	-0.51		0.33	0.76	0.79	1.36
User C	-0.87	0.33		1.00	-0.33	0.13
User D	0.19	0.76	1.00		0.94	2.89
User E	-0.19	0.79	-0.33	0.94		1.21

Calculate the rating predictions of $P_{D,3}$

$$P_{a,i} = \frac{\sum_{u=1}^n r_{u,i} \times w_{a,u}}{\sum_{u=1}^n w_{a,u}}$$

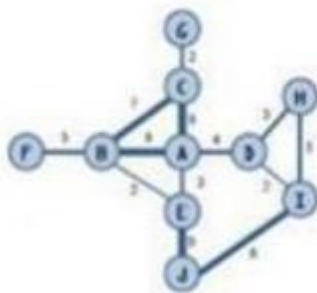
Given the ratings in the table below calculate the prediction of the Rating
P_{B,1}: 4.05

	1.-Batman Begins	2.-Alice in wonderland	3.-Dumb and Dumber	4.-Equilibrium
User A	3		3	4
User B		5	4	
User C	3	4	2	
User D	2	4		3
User E	3	4	5	

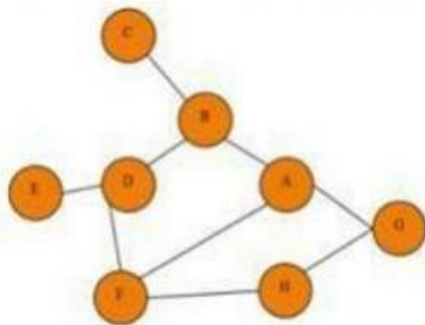
Calculate, using the formula below the prediction of the Rating **P_{B,1}**

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u)}{n}$$

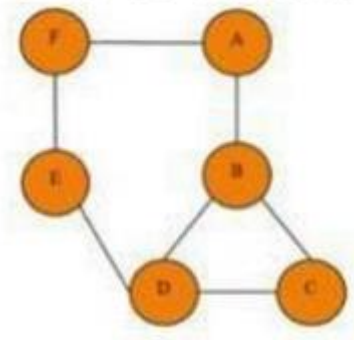
The following network is: Weighted indirect.



In the following graph, what is the most relevant node following the
 eigenvector centrality measure? A and F



What is the betweenness of node B: 2.5



Using the Playfair Cypher write -sentence: HELLO WORLD -key: SPAIN:

LCKYGTVQTKAR

S	P	A	I	N

Using the Vignère cypher give the ciphertext of “SERENDIPITY” using the key LEMON: DIDSAOMBWGJ

		--PLAINTEXT--																									
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B		B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C		C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D		D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E		E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F		F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G		G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H		H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I		I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J		J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K		K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L		L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M		M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N		N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O		O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P		P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q		Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R		R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S		S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T		T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U		U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V		V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W		W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V

Using gradient decent to optimize the least squares using the following formula, calculate θ_1 in the table below: 0.3450312

$$j = 0 : \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})$$

$$j = 1 : \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x^{(i)}$$

θ_0	0
θ_1	0.1

	X	Y			
	feet ^2	price\$	h_{θ}	$h_{\theta} - y_i$	$(h_{\theta} - y_i)^2$
	2104	460	210.40	-249.60	62300.16
	1416	232	141.60	-90.40	8172.16
	1534	315	153.40	-161.60	26114.56
	852	178	85.20	-92.80	8611.84
		suma	-594.40	105198.72	
			$J(\theta_0, \theta_1)$	13.150	

number of train	4
linear regression	$h_{\theta} = \theta_0 + \theta_1 X$

	j=0	j=1
feet ^2	$h_{\theta_1} \cdot y_i$	$(h_{\theta_1} \cdot y_i) \cdot x_i$
2104	-249.6	-525158.4
1416	-90.4	-128006.4
1534	-161.6	-247894.4
852	-92.8	-79065.6
sum/m	-148.6	-245031.2
alfa		0.000001
iter1		
θ_0		0.0001486
θ_1		

Layering and separation usually involve creating a hierarchy of visual effects, possibly matching an ordering of information content. How many layers can be seen in the pictures below: First graph 2 layers: 2 triangles, Second graph 3 layers: 1 triangle: 1 irregular pseudo-triangle, circles.



Calculate the rating predictions in the grey boxes: $P_{c4}=2.3$, $P_{d3}=5$, $P_{e4}=3$

	User A	User B	User C	User D	User E
User A	1.00	5.00	5.00	5.00	5.00
User B	5.00	1.00	5.00	5.00	5.00
User C	5.00	5.00	1.00	5.00	5.00
User D	5.00	5.00	5.00	1.00	5.00
User E	5.00	5.00	5.00	5.00	1.00

Calculate the rating predictions in the grey boxes using:

$$P_{a,i} = \frac{\sum_{u=1}^n r_{u,i} \times w_{a,u}}{\sum_{u=1}^n w_{a,u}}$$

El error reducible en una predicción se refiere o indica la dependencia de:

- Depende de la precisión de la función

La diferencia entre el análisis de texto y la búsqueda es:

- La búsqueda consiste en recuperar un documento en función de lo que los usuarios finales ya saben lo que están buscando. El análisis de texto es la información.

El correo electrónico no deseado, el reconocimiento de voz o el automóvil autónomo son ejemplos de:

- AI supervisada.

Un sistema cognitivo se refiere a:

- Un modelo de análisis de datos que responde y actúa de forma autónoma.