Feedback — Week 2 Quiz

Help Center

Thank you. Your submission for this quiz was received.

You submitted this quiz on **Wed 18 Feb 2015 5:42 AM PST**. You got a score of **9.80** out of **11.00**. You can attempt again, if you'd like.

Question 1

Suppose Coursera collected statistics on the number of students who take courses on data mining (DM) and machine learning (ML). We have the following 2×2 contingency table summarizing the statistics. If lift is used to measure the correlation between DM and ML, what is the value for lift(DM, ML)?

	DM	¬DM	Σ_{row}
ML	700	300	1000
¬ML	500	1500	2000
Σ_{col}	1200	1800	3000

Your Answer		Score	Explanation
◉ 7/4	✔	1.00	
○ -7/1200			
○ -7/4			
○ 7/1200			
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "7/4".

The lift can be calculated by

$$\text{Lift} = \frac{\text{supp}(A \cup B)}{\text{supp}(A) \cdot \text{supp}(B)},$$

where supp(A) and supp(B) refer to the relative support of A and B respectively. Thus,

$$\text{Lift} = \frac{700/3000}{1200/3000 \times 1000/3000} = 7/4.$$

Question 2

Suppose a school collected some data on students' preference for hot dogs(HD) vs. hamburgers(HM). We have the following 2×2 contingency table summarizing the statistics. If χ^2 is used to measure the correlation between HD and HM, what is the χ^2 score?

	HD	¬HD	Σ_{row}
HM	40	24	64
¬HM	210	126	336
Σ_{col}	250	150	400

Your Answer		Score	Explanation
-1			
-∞			
0	✔	1.00	
1			
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "0".

The contingency table with expected values is following table

	HD	¬HD	Σ_{row}
HM	40 (40)	24 (24)	64
¬HM	210 (210)	126 (126)	336
Σ_{col}	250	150	400

χ^2 can be evaluated as follows

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed frequency, and E_i is the expected frequency. Since the expected values equal the observed ones, we have $\chi^2 = 0$.

Question 3

What is the value range of the Kulczynski measure?

Your Answer		Score	Explanation
[-1, 1]			
(-∞, +∞)			
[0, 1]	✔	1.00	
[0, +∞)			
Total		1.00 / 1.00	

Question Explanation

By definition, the correct answer is: "[0, 1]".

Question 4

Which of the following measures is NOT null invariant?

Your Answer		Score	Explanation
Kulcyzynski			
Lift	✔	1.00	
All confidence			
Cosine			
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "Lift".

Null transactions are considered in Lift.

Question 5

Suppose we are interested in analyzing the transaction history of several supermarkets with respect to purchase of apples(A) and bananas(B). We have the following table summarizing the transactions.

Supermarket	AB	¬AB	A ¬B	¬A ¬B
S1	100,000	1,000	1,000	100
S2	50,000	7,000	3,000	600,000
S3	700,000	10,000	400,000	100,000

Which of the following measures would you use to determine the correlation of purchases between apples and bananas across all these supermarkets?

Your Answer		Score	Explanation
☑ Cosine	✔	0.25	
☐ Lift	✔	0.25	
☑ Kulcyzynski	✔	0.25	
☐ χ^2	✔	0.25	
Total		1.00 / 1.00	

Question Explanation

The correct answers are: "Kulcyzynski" and "Cosine".

Both Kulcyzynski and Cosine are null invariant measures and thus not sensitive to the null transactions, which vary dramatically for different supermarkets.

Question 6

Consider the support-based and null-invariant definitions for negative patterns. For negative pattern threshold ε = 0.01, which of the following patterns would be considered a negative pattern by the null-invariant definition but not the support-based definition?

Your Answer		Score	Explanation
○ A media content provider has 1,000,000 users. Movie A and Movie B were viewed by 1000 users each in the last month, but only 10 users viewed both.			
○ There are 5 million registered students on an online education website. 5000 students registered for Music 101, and 50,000 students registered for Data Mining, but only 500 students registered for both.			
◉ Both of the above are correct.	✖	0.00	
○ None of the above are correct.			
Total		0.00 / 1.00	

Question Explanation

The correct answer is: "A media content provider has 1,000,000 users. Movie A and Movie B were viewed by 1000 users each in the last month, but only 10 users viewed both."

Null-invariant:

$$\frac{P(A|B)+P(B|A)}{2} = \frac{10/1000+10/1000}{2} = 0.01 \leq \epsilon$$

Support based:

$$\sup(\text{Movie A}) = \sup(\text{Movie B}) = \frac{1000}{1000000} = 10^{-3}$$
$$\sup(\text{Movie A} \cup \text{Movie B}) = \frac{10}{1000000} = 10^{-5}$$
$$\sup(\text{Movie A}) * \sup(\text{Movie B}) = 10^{-3} * 10^{-3} = 10^{-6}$$
$$< \sup(\text{Movie A} \cup \text{Movie B})$$

Thus, {Movie A, Movie B} is a negative pattern by the null-invariant definition but not the support based definition.

Null-invariant:

$$\frac{P(DM|Music)+P(Music|DM)}{2} = \frac{500/5000+500/50,000}{2} = 0.055 > \epsilon$$

Support based:

$\sup(\text{DM}) = \frac{50000}{5000000} = 0.01$
$\sup(\text{Music}) = \frac{5000}{5000000} = 0.001$
$\sup(\text{DM} \cup \text{Music}) = \frac{500}{5000000} = 10^{-4}$
$\sup(\text{DM}) * \sup(\text{Music}) = 0.01 * 0.001 = 10^{-5} < \sup(\text{DM} \cup \text{Music})$

Thus, {DM, Music} is not a negative pattern by either definition.

Question 7

Pat-ID	Item-Sets	Support
P1	{A, C, E, S}	205227
P2	{F, A, C, E, S}	205211
P3	{F, A, C, E, T, S}	101758
P4	{F, A, C, T, S}	161563
P5	{A, C, T, S }	161576

Table 1: Support for frequent itemsets

Given the itemsets in Table 1, which of the following patterns are in the δ-cluster containing the pattern {A, C, E, S} for δ = 0.0001?

Your Answer		Score	Explanation
◉ {F, A, C, E, S}	✔	1.00	
○ {F, A, C, T, S}			
○ {A, C, T, S}			
○ {F, A, C, E, T, S}			
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "{F, A, C, E, S}".

All the other patterns have a distance greater than δ from the pattern {A, C, E, S}.

Question 8

Transactions	# of Transactions
(abe)	100
(bcf)	100
(acf)	100
(abcef)	100

Table 2: # transactions in a database.

Given the transactions in Table 2, which of the following is a 0.5-core pattern of (abcef)? Select all that apply

Your Answer		Score	Explanation
☑ (abe)	✔	0.20	
☐ (e)	✖	0.00	
☑ (acfe)	✔	0.20	
☑ (abcef)	✔	0.20	
☐ (a)	✔	0.20	
Total		0.80 / 1.00	

Question Explanation

The correct answers are: "(acfe)", "(abcef)", "(e)", and "(abe)".

Every pattern except (a) is contained in at most one other pattern in the database, making it a 0.5-core pattern of (abcef).

Question 9

A constraint is anti-monotone if an itemset S violates the constraint, so do all of its supersets.

Which of following constraints is anti-monotone?

Your Answer		Score	Explanation
○ var(S.price) > 20			
○ sum(S.price) > 25			
○ avg(S.profit) > 15			
◉ range(S.profit) < 10	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "range(S.profit) < 10"

Suppose X is the superset of S; if so, the following inequalities always hold for all X:

- avg(X.price) <> avg(S.price)
- sum(X.price) ≥ sum(S.price)
- var(X.profit) <> var(S.profit)
- range(X.profit) ≥ range(S.profit)

From these inequalities, we can easily verify that range(S.profit) < 10 is anti-monotone.

Question 10

A constraint is monotone if an itemset S satisfies the constraint, so do all of its supersets. Which of following constraints is NOT monotone?

Your Answer		Score	Explanation
○ v ∈ S			
○ range(S.price) > 20			
◉ avg(S.profit) < 20	✔	1.00	
○ Relative support of S < 0.1			
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "avg(S.profit) < 20"

Suppose X is the superset of S; if so, the following inequalities always hold for all X:

- Relative support of X ≤ Relative support of S
- range(X.price) ≤ range(S.price)
- avg(X.profit) <> avg(S.profit)
- v ∈ S ⊆ X

From these inequalities, we can easily verify that avg(S.profit) < 20 is not monotone.

Question 11

A constraint is succinct if the constraint c can be enforced by directly manipulating the data.

Which of following constraints is NOT succinct?

Your Answer	Score	Explanation
○ min(S.profit) < 40		
● range(S.price) > 2 ✔	1.00	It cannot be determined beforehand since range of the price of itemset S keeps increasing.
○ min(S.profit) > 40		
○ $v \in S$		
Total	1.00 / 1.00	

Question Explanation

The correct answer is: "range(S.price) > 2"