Take Test: Fall 2018 Assignment 2

Test Information

Description	
Instructions	
Multiple Attempts	Not allowed. This test can only be taken once.
Force Completion	This test can be saved and resumed later.

Question Completion Status:

QUESTION 1

5 points Saved

Consider the random experiment of picking a word from an English text article. Let W be a random variable denoting the word that we might obtain from the article. Thus W can have any value from the set of words in our vocabulary V={w_1,... ,w_N}, where w_i is a unique word in the vocabulary, and we have a probability distribution over all the words, which we can denote as {p(W=w_i)}, where p(W=w_i) is the probability that we would obtain word w_i. Now we can compute the entropy of such a variable, i.e., H(W).

Suppose we have in total N unique words in our vocabulary. What is the theoretical minimum value of H(W)?

- ○ log_2 N
- ● 0
- ○ 1
- ○ infinity

QUESTION 2

5 points Saved

[Continued from the previous question]
What is the theoretical maximum value of H(W)?

- ● log_2 N
- ○ 0
- ○ 1
- ○ infinity

QUESTION 3

5 points Saved

Suppose we have only 6 words in the vocabulary {w_1, w_2, w_3, w_4, w_5, w_6}. And there are two articles, D1 and D2 for which (1) H(W) achieves the minimum value in D1, and the maximum value in D2, respectively; (2) the two articles contain the same number of words in total. How many w_1 are there in D1 and D2? **Select all possible cases that apply**:

- [x] 6 in D1, and 1 in D2
- [ ] 1 in D1, and 6 in D2
- [ ] 1 in D1, and 0 in D2
- [x] 0 in D1, and 1 in D2
- [ ] 0 in D1, and 6 in D2
- [ ] 1 in D1, and 1 in D2

Click Save and Submit to save and submit. Click Save All Answers to save all answers.

Save All Answers Save and

Suppose we have a vocabulary set V = {w_1, w_2, w_3, w_4}. Each of articles D_1, D_2, …, D_4 contain only words from V, and H(W) = 0 for all 4 articles. Suppose we concatenate D_1 through D_4 to form a longer article D_5. What is the maximum possible value of H(W) in article D_5? (Assuming the base of log is 2).

- ○ 0.5
- ○ 1
- ◉ 2
- ○ 4

QUESTION 5

5 points Saved

Suppose we have a vocabulary set V = {w_1, w_2, w_3, w_4}. Each of articles D_1, D_2, …, D_4 contain only words from V, and H(W) = 2 for all 4 articles. Suppose we concatenate D_1, D_2 through D_4 to form a longer article D_5. What is the maximum possible value of H(W) in article D_5? Assuming the base of log is 2.

- ○ 0.5
- ○ 1
- ◉ 2
- ○ 4

QUESTION 6

5 points Saved

What is the value of the conditional entropy H(X I X)?

- ◉ 0
- ○ 1
- ○ H(X)
- ○ log p(X|X)

QUESTION 7

5 points Saved

What is the value of mutual information I(X;Y) if X and Y are independent?

- ○ log_2 N
- ◉ 0
- ○ H(X|Y)

QUESTION 8

5 points Saved

Mutual information can be used to measure the correlation of two words. Suppose we have a collection of N documents. For a word A in the collection, we use p(X_A), where X_A \in {0, 1}, to represent the probability that A occurs (X_A=1) in one document or not (X_A=0). If word A appears in N_A documents, then p(X_A=1) = N_A / N and p(X_A=0) = (N-N_A) / N. Similarly, we can define the probability p(X_B) for another word B. We also define the joint probability of word A and B as follows:

p(X_A =1,X_B =1): the probability of word A and word B co-occurring in one document. If there are N_{AB} documents containing both word A and B in the collection, then p(X_A=1, X_B=1) = N_{AB} / N

p(X_A =1,X_B =0): the probability that word A occurs in one document but B does not occur in that document. It can be calculated as p(X_A=1, X_B=0) =(N_A - N_{AB}) / N
Write down the formulas for p(X_A=0, X_B=1) and p(X_A=0, X_B=0). Select the two formulas from the following options (**select all that applies**):

- ☐ (N_A - N_{AB}) / N
- ☑ (N_B - N_{AB}) / N
- ☐ (N - N_A - N_B - N_{AB}) / N
- ☑ (N - N_A - N_B + N_{AB}) / N

Click Save and Submit to save and submit. Click Save All Answers to save all answers.

Save All Answers Save and

Next, we will use the following tables to do some real computation of Mutual Information. The tables contain the document counts for different words. There are a total of N = 26,394 documents in the collection.
Table 1 contains the document counts for words ‘computer’ and ‘program’, derived from the document collection (Hint: If A = computer and B = program, then N_{AB} = 349. This means there are 349 documents that contain ‘computer’ AND ‘program’):

	p_{computer} = 1	p_{computer} = 0
p_{program} = 1	349	2021
p_{program} = 0	1041	22983

(Table 1: "computer" and "program")

Table 2 contains the document counts for words ‘computer’ and ‘baseball’, derived from the same document collection:

	p_{computer} = 1	p_{computer} = 0
p_{baseball} = 1	23	2121
p_{baseball} = 0	1367	22883

(Table 2: "computer" and "baseball")

Use the document counts from Table 1 and 2 to compute the value of x = I(X_{computer}; X_{program}).
Fill in the following blank with the value of x:
(Requirement for filling the blanks: you must round the result to **4 decimal places** for exact match of the answer, i.e., **you won't get any credit for rounding to 2 decimals or 3 decimals places even your answer is correct**.
For example, 0.3333333 => 0.3333; 0.6666666 => 0.6667)
(Assume the base of log is 2.)
[x]

0.0092

QUESTION 10

5 points Saved

[Continued from the previous question]
Use the document counts from Table 1 and 2 from the previous question to compute y = I(X_{computer}; X_{baseball}). Fill in the following blank with the value of y:
(Requirement for filling the blanks: you must round the result to **4 decimal places** for exact match of the answer, i.e., **you won't get any credit for rounding to other number of decimals places even your answer is correct**.
For example, 0.3333333 => 0.3333 (correct), 0.33 (wrong); 0.6666666 => 0.6667 (correct), 0.666667 (wrong))
(Assume the base of log is 2.)
[y]

0.0032

QUESTION 11

5 points Saved

(Continued from the previous question)
Using Table 1 and Table 2 from the previous question, compute the values of z = I(X_{program}; X_{computer}) and w = I(X_{baseball}; X_{computer}). Fill in the following two blanks with the values of z:
(Requirement for filling the blanks: you must round the result to **4 decimal places** for exact match of the answer, i.e., **you won't get any credit for rounding to other number of decimals places even your answer is correct**.
For example, 0.3333333 => 0.3333 (correct), 0.33 (wrong); 0.6666666 => 0.6667 (correct), 0.666667 (wrong))
(Assume the base of log is 2.)
z: 0.0092 ,

and w: 0.0032 .

QUESTION 12

5 points Saved

Compare the results of I(X_{computer};X_{program}) and I(X_{computer};X_{baseball}).

Click Save and Submit to save and submit. Click Save All Answers to save all answers.

Save All Answers Save and

QUESTION 13

5 points Saved

[Continued from the previous question]
Do the comparative results from the previous question conform with your intuition? Explain your intuition.

For the toolbar, press ALT+F10 (PC) or ALT+FN+F10 (Mac).

Paragraph Arial 3 (12pt)

Mashups

It conforms with my intuition that 'program' and 'computer' gives more mutual information. 'Program' and 'computer' have more correlation than 'baseball' and 'computer' since there are more documents that contain both 'computer' and 'program' and neither 'computer' nor 'program' (349+22983 = 23332 > 23+22883 = 22906).

Path: p Words:42

QUESTION 14

10 points Saved

Whats the range of KL Divergence?

- ○ (-infinity, 0]
- ○ [0, infinity)
- ◉ [0, log_2 N)
- ○ (0, log_2 N)

QUESTION 15

10 points Saved

[Continued from the previous question]

Under which circumstances does D_{KL}(p||q) equal to 0?

- ◉ p shares the same distribution as q but they can be any distribution
- ○ p shares the same distribution as q, and the shared distribution can only be the uniform distribution

QUESTION 16

5 points Saved

Given Table 1 from question 3b (i.e., "computer" and "program"), compute the KL divergence between p(X_{computer}) and p(X_{computer}|X_{program} = 1):

Compute the values of x and y:

x = D_{KL}(p(X_{computer}) || p(X_{computer}|X_{program} = 1)) and

y = D_{KL}(p(X_{computer}|X_{program} = 1) || p_{computer})

Fill in the following blank with the values of x:

Does x equals to y?

(Requirement for filling the blanks: you must round the result to **4 decimal places** for exact match of the answer, i.e., **you won't get any credit for rounding to 2 decimals or 3 decimals places even your answer is correct**.

For example, 0.3333333 => 0.3333; 0.6666666 => 0.6667)

x: 0.0656 ,

and y: 0.0890

Note: the log base of KL divergence is 2, throughout question 4a-4c

QUESTION 17

5 points Saved

When calculating D_{KL} (p||q) for two probability distributions p and q, if there exists some values i such that p(i) = 0 or q(i) = 0, can you still compute D_{KL} (p||q)? i.e., the output D_{KL}(p||q) is a valid value that

Click Save and Submit to save and submit. Click Save All Answers to save all answers.

Save All Answers Save and

- [x] p(i): yes, even there exist i such that p(i) =0, the output D_{KL} (p|| q) is still valid;
  q(i): no, if there exists i such that q(i) =0, the output D_{KL}(p||q) is not valid
- [ ] p(i): no, if there exists i such that p(i) =0, the output D_{KL} is not valid;
  q(i): yes, even there exist i such that q(i) =0, the output D_{KL} (p|| q) is still valid;
- [ ] p(i): no, if there exists i such that p(i) =0, the output D_{KL}(p||q) is not valid;
  q(i): no, if there exists i such that q(i) =0, the output D_{KL}(p||q) is not valid;

QUESTION 18

5 points Saved

Suppose there are two distributions p and q, and the two distributions are in Table 3 as below.

i	p	q
0	0.5	0.5
1	0.5	0
2	0	0.5

Table 3

Here some values of p and q are 0. We can resolve the 0 probabilities issue using the smoothing technique in information retrieval.

To smooth p and q, one way is to replace p(i) in Table 3 with a smoothing function, (1-lambda) p(i) + lambda * 1/N, where N is the number of unique values of p, i.e., N = 3 in Table 3. Similarly, we can replace q(i) in Table 3 with (1-lambda) q(i) + lambda * 1/N. After smoothing **for both p and q**, What is the new value of x = D_{KL} (p||q) when lambda = 0.1? Fill in the blank below with the value of x:

(Requirement for filling the blanks: you must round the result to **4 decimal places** for exact match of the answer, i.e., **you won't get any credit for rounding to 2 decimals or 3 decimals places even your answer is correct**.
For example, 0.3333333 => 0.3333; 0.6666666 => 0.6667)
[x]

1.7361

Click Save and Submit to save and submit. Click Save All Answers to save all answers.

Save All Answers Save and