

# ECS763U/ECS763P - Natural Language Processing - 2022/23 - Semester 1

Started on	Sunday, 23 October 2022, 11:48 PM
State	Finished
Completed on	Monday, 24 October 2022, 12:18 AM
Time taken	30 mins 28 secs
Grade	10.00 out of 10.00 (100%)

Question 1

Correct

Mark 1.00 out of 1.00

Flag question

Select which of the following is true for training a classifier.

Select one:

- ☐ a. Ideally, the number of features should be considerably greater than the number of instances of training data.
- ☒ b. Ideally, the number of instances of training data should be considerably greater than the number of features.

The correct answer is: Ideally, the number of instances of training data should be considerably greater than the number of features.

Question 2

Correct

Mark 1.00 out of 1.00

Flag question

The conjunctive probability of the event of both A and B being the case (i.e.  $p(A \wedge B)$ ) can be the same as the probability of either of the individual outcomes on their own (i.e. the same as  $p(A)$  or the same as  $p(B)$ ).

Select one:

- ☒ True
- ☐ False

The correct answer is 'True'.

Question 3

Correct

Mark 1.00 out of 1.00

Flag question

The conjunctive probability of the event of both A and B being the case (i.e.  $p(A \wedge B)$ ) can be less than the probability of each of the individual outcomes on their own (i.e. less than  $p(A)$  and less than  $p(B)$ ).

Select one:

- ☒ True
- ☐ False

The correct answer is 'True'.

Question 4

Correct

Mark 1.00 out of 1.00

Flag question

Which of these are valid probability values (select all that apply)?

Select one or more:

- ☒ a. 0
- ☒ b. 1
- ☐ c. 1.01
- ☐ d. -0.01
- ☒ e. 1.00
- ☐ f. 1.3

The correct answers are: 0, 1.00, 1

Question 5

Correct

Mark 1.00 out of 1.00

Flag question

Which stage of supervised text classification comes first?

Select one:

- ☐ a. Tokenization.
- ☐ b. Stemming.
- ☐ c. Pre-processing text.
- ☒ d. Reading in the text data to the programme.
- ☐ e. Feature extraction

The correct answer is: Reading in the text data to the programme.

Question 6

Correct

Mark 1.00 out of 1.00

Flag question

If the likelihood of outcome B,  $p(B)$ , is 0.2 and the likelihood of  $p(A|B)$  is 0.8, what is the likelihood  $p(A \wedge B)$ ?

Select one:

- ☒ a. 0.16
- ☐ b. Not enough information given to know.
- ☐ c. 0.6
- ☐ d. 0.25
- ☐ e. 1.0

The correct answer is: 0.16

Question 7

Correct

Mark 1.00 out of 1.00

Flag question

A Bayesian classifier is one which uses the  probability, the likelihood, and the evidence to calculate the posterior probability.

The correct answer is: prior

Question 8

Correct

Mark 1.00 out of 1.00

Flag question

Where P is Precision and R is recall, fill in the missing part of the equation below for the formula for the F1 score of a classifier's predictions. F1 =  \* PR / (P + R)

The correct answer is: 2

Question 9

Correct

Mark 1.00 out of 1.00

Flag question

Consider the table of word counts for a small vocabulary of 7 words over a corpus of positive and negative tweets.

	#DOCS	bieber	love	hate	#sarcasm	hi	lol	nlp	#Total words in docs
POSITIVE	400	40	30	23	40	120	300	30	583
NEGATIVE	100	12	4	20	38	27	1	3	105

$$C_{NB} = \operatorname{argmax}_{c_j \in C} \log(P(c_j)) + \sum_{i \in \text{positions}} \log(P(x_i | c_j))$$

	$\log(P(c_j))$	$\sum_{i \in \text{positions}} \log(P(x_i   c_j))$	$\log(P(c_j)) + \sum_{i \in \text{positions}} \log(P(x_i   c_j))$
POSITIVE			
NEGATIVE	?		

You are applying a Naive Bayes Classifier trained on this data to the tweet "**love bieber #sarcasm**". What is the value for the part of the equation indicated by the ? in the table- i.e. the log of the prior for the class NEGATIVE? Use log base 2 and give your answer to 2 DECIMAL PLACES.

Answer:

Yes, correct for log base 2.

The correct answer is: -1.61

Question 10

Correct

Mark 1.00 out of 1.00

Flag question

Consider the table of word counts for a small vocabulary of 7 words over a corpus of positive and negative tweets.

	#DOCS	bieber	love	hate	#sarcasm	hi	lol	nlp	#Total words in docs
POSITIVE	400	40	30	23	40	120	300	30	583
NEGATIVE	100	12	4	20	38	27	1	3	105

$$C_{NB} = \operatorname{argmax}_{c_j \in C} \log(P(c_j)) + \sum_{i \in \text{positions}} \log(P(x_i | c_j))$$

	$\log(P(c_j))$	$\sum_{i \in \text{positions}} \log(P(x_i   c_j))$	$\log(P(c_j)) + \sum_{i \in \text{positions}} \log(P(x_i   c_j))$
POSITIVE		?	
NEGATIVE			

You are applying a Naive Bayes Classifier trained on this data to the tweet "**love bieber #sarcasm**". What is the value for the part of the equation indicated by the ? in the table- i.e. the sum of the logs of the likelihood of each word in the tweet given the class POSITIVE? Assume that the word counts have already had 1 added to them already in smoothing. Use log base e and give your answer to 2 DECIMAL PLACES.

Answer:

Yes, correct for base e.

The correct answer is: -8.33

Finish review

Lecture live Zoom link for remote participants

Jump to...

Unit 3 Slides

## Student Life

Student email  
My QMUL  
Queen Mary Students' Union  
Student Enquiry Centre  
QMplus for students  
Careers

## Library

Library Landing Page  
Library Website  
Find It! Use It! Reference It!  
Library Search  
Subject Guides  
Cite Them Right  
Academic Skills

## Archives

Archive  
2021/22  
2020/21  
2019/20  
2018/19  
2017/18