Feedback — Week 3 Practice Quiz

Help Center

Thank you. Your submission for this quiz was received.

You submitted this quiz on **Sun 28 Jun 2015 12:13 PM PDT**. You got a score of **11.00** out of **11.00**.

Question 1

You are given two unigram language models θ_1 and θ_2 as defined in the table below:

w	$P(w\|\theta_1)$	$P(w\|\theta_2)$
the	0.4	0.05
of	0.4	0.05
technology	0.1	0.5
machine	0.1	0.4

Suppose we are using a mixture model for document clustering based on the two given unigram language models, θ_1 and θ_2, such that $P(\theta_1) = 0.3$ and $P(\theta_2) = 0.7$. To generate a document, first, one of the two language models is chosen according to $P(\theta_i)$, and then all the words in the document are generated based on the chosen language model.

The probability of generating a document composed only of the one word "technology" using the given mixture model is $P(\text{“technology”}) =$

Your Answer		Score	Explanation
◉ 0.38	✔	1.00	
○ 0.3			
○ 0.6			
○ 0.58			

Total	1.00 / 1.00

Question Explanation

$P(\text{“technology”}) = 0.3 * 0.1 + 0.7 * 0.5 = 0.38$

Question 2

Assume the same given as in Question 1. What is the probability of generating a document composed only of the phrase “the technology,” i.e., $P(\text{“the technology”})$?

Your Answer		Score	Explanation
0.3			
0.0295	✔	1.00	
0.0589			
0.1444			
Total		1.00 / 1.00	

Question Explanation

$P(\text{“the technology”}) = 0.3 * 0.4 * 0.1 + 0.7 * 0.05 * 0.5 = 0.0295$

Question 3

Suppose we are performing document clustering on a collection of N documents using a mixture model as discussed in the lecture **Text Clustering: Generative Probabilistic Models (Part 3)**. Let the number of clusters be K and the vocabulary size be M. What is the number of parameters that the EM algorithm tries to estimate? Consider each $P(\theta_i)$ or $P(w|\theta_i)$ as a separate parameter.

Your Answer		Score	Explanation
$K + MK$	✔	1.00	
$KN + MK$			
MNK			
MK			

Total	1.00 / 1.00

Question Explanation

K parameters are needed to estimate $P(\theta_i)$ for each $i \in \{1, \ldots, K\}$.

MK parameters are needed to estimate $P(w|\theta_i)$ for each $w \in V$ and $i \in \{1, \ldots, K\}$.

Question 4

Assume that documents are being classified into two categories, c1 and c2, such that a document can belong to more than one category. The table below shows the prediction of a classifier, denoted by “y” or “n", in addition to the true label (ground truth) represented by a “+” or “-”, where a correct prediction is either y (+) or n (-).

	c1	c2
D1	y(+)	y(+)
D2	n(-)	y(+)
D3	n(+)	n(-)
D4	y(-)	y(+)
D5	n(+)	n(-)

Let P(ci) and R(ci) denote the precision and recall associated with category ci, respectively.

The precision and recall of c1 and c2 are:

Your Answer					Score	Explanation
○ P(c1) = 1/2	R(c1) = 1/2	P(c2) = 1	R(c2) = 1			
◉ P(c1) = 1/2	R(c1) = 1/3	P(c2) = 1	R(c2) = 1	✔	1.00	
○ P(c1) = 1/3	R(c1) = 1/2	P(c2) = 1	R(c2) = 1			
○ P(c1) = 1/2	R(c1) = 1/2	P(c2) = 1/2	R(c2) = 1/2			
Total					1.00 / 1.00	

Question Explanation

P(ci) = TP(ci)/(TP(ci) + FP(ci)) and R(ci) = TP(ci)/(TP(ci)+FN(ci))

Question 5

Given the same data as in Question 4, the classification accuracy of the classifier is:

Your Answer		Score	Explanation
9/10			
8/10			
7/10	✔	1.00	
3/10			
Total		1.00 / 1.00	

Question Explanation

Classification accuracy is the ratio of the number of correct decisions (i.e., y (+) and n(-)) to the total number of decisions made.

Question 6

Given the same data as in Question 4, what is the recall of the classifier using **micro-averaging** (i.e., by pooling all decisions together)?

Your Answer		Score	Explanation
1			
2/3	✔	1.00	
4/5			
5/6			
Total		1.00 / 1.00	

Question Explanation

Consider all the predictions “pooled” together, then R = [number of y(+)]/[number of y(+) + number of n(+)] = 4/6 = 2/3.

Question 7

The following table shows the **similarity** values between a set of documents as well as a binary label associated with each document.

	D1	D2	D3	D4	Label
D1	1.0	0.05	0.9	0.8	1
D2	0.05	1.0	0.01	0.5	0
D3	0.9	0.01	1.0	0.7	1
D4	0.8	0.5	0.7	1.0	?

Suppose we use {D1,D2,D3} as our training dataset and use the k-Nearest Neighbor classifier to predict the label of D4. If k=1, then the prediction of the classifier for D4 is:

Your Answer		Score	Explanation
◉ 1	✔	1.00	
○ 0			
Total		1.00 / 1.00	

Question Explanation

D1 is the most similar document to D4, and the label of D1 is 1, thus the nearest neighbor classifier will predict 1.

Question 8

Assume the same given as in Question 7. Then, the prediction for D4 whether we use k=2 or k=3 will be 1.

Your Answer		Score	Explanation
○ False			
◉ True	✔	1.00	

Total	1.00 / 1.00

Question Explanation

In both cases, the majority of votes will be for the label 1.

Question 9

To apply Naive Bayes Classification, we first need to estimate the parameters $P(w|\theta_i)$ and $P(\theta_i)$ for each corresponding category i. Suppose we would like to do binary classification. Consider the following corpus of two documents, d_1 and d_2 associated with two categories, T_1 and T_2. Each category contains one document as follows:

$$T_1 : \{d_1 = (w_1 w_1 w_1 w_1 w_3 w_3)\}$$
$$T_2 : \{d_2 = (w_1 w_1 w_2 w_2 w_3 w_4)\}$$

We estimate the parameters using the maximum likelihood estimator, i.e., $P(w|\theta_i) = \frac{c(w,T_i)}{|T_i|}$ and $P(\theta_i) = \frac{|T_i|}{\sum_j |T_j|}$, where $|T_i|$ is the total number of words in category i.

Given a new document $d_3 = (w_3, w_4)$, what will $P(d_3|\theta_1)$ be?

Your Answer		Score	Explanation
0	✔	1.00	
0.25			
0.5			
1			
Total		1.00 / 1.00	

Question Explanation

$P(d_3|\theta_1) = P(w_3|\theta_1)P(w_4|\theta_1) = 0$ since $P(w_4|\theta_1) \propto c(w_4, T_1) = 0$.

Question 10

Suppose that we now use Laplace smoothing, what is $P(\theta_1|d_3)$?

Note that Laplace smoothing is an additive smoothing method that is defined by $P(w|\theta_i) = (c(w,T_i)+1)/(|T_i|+|V|)$ where |V| is the size of the vocabulary in the training data (i.e., the number of unique terms in the training data).

Your Answer		Score	Explanation
3/7	✔	1.00	
2/29			
5/8			
1/2			
Total		1.00 / 1.00	

Question Explanation

$$P(\theta_1|d_3) \propto P(d_3|\theta_1)P(\theta_1)$$
$$= P(w_3|\theta_1)P(w_4|\theta_1)P(\theta_1)$$
$$= \frac{c(w_3,T_1)+1}{|T_1|+|V|}\frac{c(w_4,T_1)+1}{|T_1|+|V|} * 1/2$$
$$= (\frac{2+1}{4+6})\frac{0+1}{4+6} * 1/2$$
$$= 3/10 * 1/10 * 1/2$$
$$= (3/2)/100$$

$$P(\theta_2|d_3) \propto P(d_3|\theta_2)P(\theta_2)$$
$$= P(w_3|\theta_2)P(w_4|\theta_2)^2P(\theta_2)$$
$$\frac{c(w_3,T_2)+1}{|T_2|+|V|}\frac{c(w_4,T_2)+1}{|T_2|+|V|}$$
$$\frac{1+1}{4+6}(\frac{1+1}{4+6}) * 1/2$$
$$2/10 * 2/10 * 1/2$$
$$2/100$$

Thus, $P(\theta_1|d_3) = \frac{(3/2)/100}{(3/2)/100+2/100}$

$$= \frac{3/2}{3/2+2}$$
$$= \frac{3}{3+4}$$
$$= \frac{3}{7}$$

Question 11

Which category would Naive Bayes predict for d_3 if we use Laplace smoothing?

Your Answer		Score	Explanation
◉ Category 2	✔	1.00	
○ Category 1			
Total		1.00 / 1.00	

Question Explanation

Since $P(\theta_1|d_3) + P(\theta_2|d_3) = 1$, and $P(\theta_1|d_3) = 3/7$, we have $P(\theta_2|d_3) = 4/7$.