Feedback — Week 3 Quiz

Help Center

You submitted this quiz on **Sat 11 Jul 2015 11:05 AM PDT**. You got a score of **10.00** out of **10.00**.

Question 1

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

w	$P(w\|\theta_1)$	$P(w\|\theta_2)$
concert	0.1	0.4
music	0.1	0.4
data	0.4	0.1
software	0.4	0.1

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 $P(\theta_1) = 0.5$ and $P(\theta_2) = 0.5$. 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 the document d: "music software" using the given mixture model is $P(\text{“music software”}) =$

Your Answer		Score	Explanation
0.05			
0.0625			
0.5			
0.04	✔	1.00	
Total		1.00 / 1.00	

Question 2

Assume the same unigram language models θ_1 and θ_2 defined as in the table of Question 1 with $P(\theta_1) = 0.5$ and $P(\theta_2) = 0.5$.

We now want to generate documents based on the mixture model used in topic modeling. To generate a document, **for each word**, we first choose one of the two language models, θ_1 and θ_2, and then generate the word according to the chosen model. The probability of generating the document d: "music software" according to this mixture model is $P(\text{"music software"}) =$

Your Answer		Score	Explanation
0.5			
0.05			
0.0625	✔	1.00	
0.04			
Total		1.00 / 1.00	

Question 3

Suppose we have the following training dataset of emails where each email is associated with the label spam or ham (not-spam). We want to train a Naive Bayes classifier based on this dataset.

	Document	Words in Document	Spam/Ham
Training Data	d_1	Save Money No Fees	Spam
	d_2	Back to the Future	Ham
	d_3	Back to School Night	Ham

Using maximum likelihood estimation without smoothing, what is $P(\text{Spam})$?

Your Answer		Score	Explanation
1/3	✔	1.00	

1/2	
1/4	
1/5	
Total	1.00 / 1.00

Question 4

Assume the same given as in Question 3 and that additive probability smoothing is being used to evaluate $P(w|\text{Spam})$ and $P(w|\text{Ham})$, i.e., $P(w|\text{Spam}) = \frac{c(w,\text{Spam})+1}{\sum_{w'} c(w',\text{Spam})+|V|}$ and $P(w|\text{Ham}) = \frac{c(w,\text{Ham})+1}{\sum_{w'} c(w',\text{Ham})+|V|}$ where $|V| = 10$ is the size of the vocabulary.

Which of the following documents has the **highest** probability of being classified as **spam** by the Naive Bayes classifier?

Hint: You should not need to compute the actual probabilities to answer this question. You can answer it by inspecting the score function on the slide entitled "Anatomy of Naïve Bayes Classifier."

Your Answer		Score	Explanation
“No fees”	✔	1.00	
“Save money back”			
“Save money future”			
“Future school no fees”			
Total		1.00 / 1.00	

Question 5

The following table shows the **similarity** values between a set of emails as well as a binary label associated with each email indicating whether it is spam (label=1) or ham (label=0).

	D1	D2	D3	D4	D5	D6	Label

D1	1.0	0.1	0.5	0.8	0.82	0.85	1
D2	0.1	1.0	0.85	0.05	0.12	0.7	0
D3	0.5	0.85	1.0	0.1	0.1	0.6	0
D4	0.8	0.05	0.5	1.0	0.9	0.1	1
D5	0.82	0.12	0.1	0.9	1.0	0.3	1
D6	0.85	0.7	0.6	0.1	0.3	1.0	?

Suppose we use {D1,D2,D3,D4,D5} as our training dataset and use the k-Nearest Neighbor classifier to predict the label of email D6. If k=1, then the prediction of the classifier for D6 is:

Your Answer		Score	Explanation
◉ 1	✔	1.00	
○ Cannot be decided			
○ 0			
Total		1.00 / 1.00	

Question 6

Assume the same setup as in Question 5. Moreover, in case of equal number of votes for both labels, assume that the predicted label will be 1. For which of the following values of k will the prediction of the classifier for D6 be 0?

Your Answer		Score	Explanation
○ 4			
○ 2			
◉ 3	✔	1.00	
○ 5			
Total		1.00 / 1.00	

Question 7

Which of the following is **not** true?

Your Answer		Score	Explanation
◉ K-NN tries to estimate d+1 weights associated with d features.	✔	1.00	
○ SVM and Logistic Regression try to estimate d+1 weights associated with d features.			
○ A linear SVM tries to maximize the margin between the linear separator and the two categories of the training data.			
○ Naive Bayes is a generative classifier while K-NN is discriminative.			
Total		1.00 / 1.00	

Question 8

Suppose we are performing clustering on a collection of documents using a mixture model as discussed in the lecture **Text Clustering: Generative Probabilistic Models (Part 3)**. Then, if we add more documents to the collection such that no new words are added to the vocabulary, the number of parameters to be estimated by the EM algorithm, i.e., $P(\theta_i)$ and $P(w|\theta_i)$, will:

Note: Do **not** count the probabilities associated with the hidden variables (i.e., those estimated in the E-step) as parameters.

Your Answer		Score	Explanation
◉ Stay the same	✔	1.00	
○ Increase			
○ Decrease			
Total		1.00 / 1.00	

Question 9

Assume that documents are being classified into 3 categories, c1, c2, and c3 such that a document can belong to multiple categories. 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	c3
D1	y(+)	y(-)	n(+)
D2	n(-)	y(+)	n(-)
D3	y(+)	n(-)	y(+)
D4	y(+)	y(+)	y(+)

Let P(ci), R(ci), and F(ci) denote the precision, recall, and F1 measure associated with category ci, respectively.

Which of the following is **not** true?

Your Answer		Score	Explanation
○ P(c1) = 1 R(c1) = 1			
○ P(c2) = 2/3 R(c2) = 1			
◉ P(c3) = 2/3 R(c3) = 1	✔	1.00	
○ F(c2) = F(c3) = 4/5			
Total		1.00 / 1.00	

Question 10

Given the same data as in Question 9, what are the **precision** and **recall** values of the classifier using **micro-averaging** (i.e., by pooling all decisions together)?

Your Answer	Score	Explanation

○ P = 1 R = 1		
○ P = 7/12 R = 8/12		
○ P = 7/12 R = 7/12		
◉ P = 7/8 R = 7/8	✔	1.00
Total		1.00 / 1.00