# Week 5 Practice Quiz

**10/10** points earned (100%)

Excellent!

Retake

[Course Home](https://www.coursera.org/learn/text-mining/home/welcome)

Correct

1 / 1 points

1. Sentiment classification can be treated as a text categorization problem.

1. **True**

**Correct Response**

Sentiment classification is to classify a piece of text into positive, negative (or neutral)

1. False

Correct

1 / 1 points

2. Suppose we want to perform rating prediction using 10 features where the output of the classifier is a discrete prediction r∈{1,2,3,4,5}. Then, the number of parameters that need to be estimated using ordinal logistic regression is **less** than the number of parameters that need to be estimated if we use 4 independent logistic regression classifiers to perform prediction.

1. **True**

**Correct Response**

Ordinal logistic regression shares the parameters

1. False

Correct

1 / 1 points

3. 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?

1. P(c3) = 2/3 R(c3) = 1
2. P(c1) = 1 R(c1) = 1
3. **F(c2) = F(c3) = 4/5**

**Correct Response**

See slides for the formula of F-measure

Correct

1 / 1 points

4. Given the same data as in Question 3, what are the **precision** and **recall** values of the classifier using **micro-averaging** (i.e., by pooling all decisions together)?

1. **P = 7/8 R = 7/8**

**Correct Response**

micro-average is computed by pooling all decisions together

1. P = 7/12 R = 7/12
2. P = 7/12 R = 8/12
3. P = 1 R = 1

Correct

1 / 1 points

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

1. A linear SVM tries to maximize the margin between the linear separator and the two categories of the training data.
2. Naive Bayes is a generative classifier while K-NN is discriminative.
3. SVM and Logistic Regression try to estimate d+1 weights associated with d features.
4. **K-NN tries to estimate d+1 weights associated with d features.**

**Correct Response**

k-NN only computes for each data which one is its nearest neighbor and there is no weights to estimate.

Correct

1 / 1 points

6. 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.

* d1 is Spam and have words: Save Money No Fees
* d2 is Ham and have words: Back to the Future
* d3 is Ham and have words: Back to School Night

Using maximum likelihood estimation without smoothing, what is P(Spam)?

1. **1/3**

**Correct Response**

the prior probability of spam is 1 (d1) out of 3 (d1, d2, d3)

1. 1/2
2. 1/4
3. 1/5

Correct

1 / 1 points

7. Assume the same given as in Question 6 and that additive probability smoothing is being used to evaluate P(w|Spam) and P(w|Ham), i.e., *P*(*w*|*Spam*)=*c*(*w*,*Spam*)+1∑*w'c*(*w'*,*Spam*)+|*V*| and *P*(*w*|*Ham*)=*c*(*w*,*Ham*)+1∑*w'c*(*w'*,*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."

1. **“No fees”**

**Correct Response**

No and Fee only show up in Spam

1. “Save money back”
2. “Save money future”
3. “Future school no fees”

Correct

1 / 1 points

8. To apply Naive Bayes classification, we first need to estimate the parameters P(w|θi) and P(θi) for each corresponding category i. Suppose we would like to do binary classification. Consider the following corpus of two documents, d1 and d2 associated with two categories, T1 and T2. Each category contains one document as follows:

* T1:{d1=(w1w1w1w1w3w3)}
* T2:{d2=(w1w1w2w2w3w4)}

We estimate the parameters using the maximum likelihood estimator, i.e., P(w|θi)=c(w,Ti)|Ti| and P(θi)=|Ti|∑j|Tj|, where |Ti| is the total number of words in category i.

Given a new document d3=(w3,w4), what will P(d3|θ1) be?

1. **0**

**Correct Response**

*P*(*w*4|*θ*1)=0

1. 1
2. 0.6
3. 0.5

Correct

1 / 1 points

9. Suppose that we now use Laplace smoothing, what is P(θ1|d3)?

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

1. 1/2
2. **3/7**

**Correct Response**

(1/10 \* 3/ 10) / ( (1/10 \* 3/ 10) + (2/10 \* 2/ 10) )

1. 5/8
2. 2/29

Correct

1 / 1 points

10. Which category would Naive Bayes predict for d3 if we use Laplace smoothing?

1. Category 1
2. **Category 2**

**Correct Response**

the posterior probability for category 2 is 4/7