

Q8

Due Nov 21 at 11:59pm **Points** 1 **Questions** 10
Available after Nov 18 at 1:59pm **Time Limit** 20 Minutes

Instructions

This quiz covers material from: **week 8**

Reminders:

- There are 10 questions.
- You have only one attempt.
- You have 20 minutes.
- All lecture quizzes must be completed by Monday of the following week.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	20 minutes	1 out of 1

❗ Correct answers are hidden.

Score for this quiz: **1** out of 1

Submitted Nov 21 at 5:07pm

This attempt took 20 minutes.

Question 1

0.1 / 0.1 pts

In the text+ml notebook, we created TF-IDF vectors using sklearn like this

```
# Learn vocabulary and idf, return term-document matrix.
# return an array; our predictor
tfidf_X = tfidf.fit_transform(df_train['response']).toarray()

# take a look at the output
print(tfidf_X.shape)

print("min: ", np.min(tfidf_X), '\n',
      "mean: ", np.mean(tfidf_X), '\n',
      "max: ", np.max(tfidf_X))

(2993, 500)
min: 0.0
mean: 0.006827444081791176
max: 1.0
```

select all of the following statements which are true



.fit_transform() creates the TF-IDF vector from the student response text



there are 2993 unique words (vocabulary) in the student responses



there are 500 student responses



there are 2993 student responses



there are 500 unique words (vocabulary) in the student responses

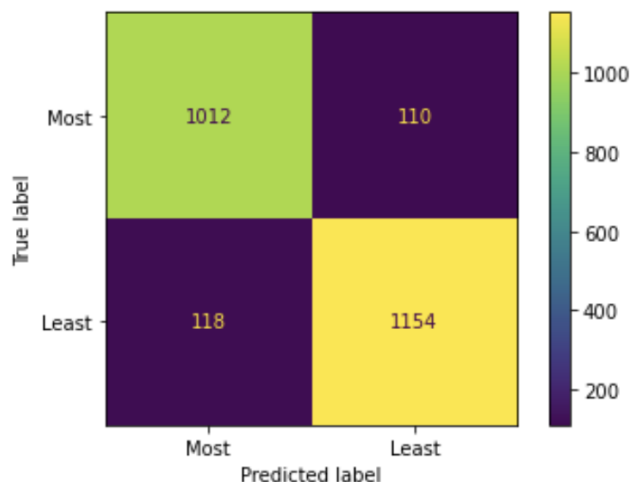
Question 2

0.1 / 0.1 pts

Here is a confusion matrix we generated from using a Support Vector Classifier to try to predict whether a student's response was a comment about something they "most liked" or "least liked" about the class.

!]:

```
cm = confusion_matrix(tfidf_train_Y, df_predicted_train_Y)
disp = ConfusionMatrixDisplay(cm, ['Most', 'Least'])
disp.plot();
```



Let's assume that "Least liked" is the positive class such that the lower right square of the matrix shows the True Positive count and the upper left square shows the True Negative count.

Given the above, where does one find the False Positives?

- ☐ Bottom Left
- ☒ Top Right
- ☐ You can't figure it out from this matrix
- ☐ By adding the numbers of the bottom row

Question 3

0.1 / 0.1 pts

A Support Vector Machine is great for classification because it

- ☐ predicts the label of each datapoint using the label from the nearest neighbor in the training set
- ☒ finds the decision boundary that separates two classes such that there is the BIGGEST space (margin) between the boundary and the nearest training samples (support vectors).
- ☐ finds the mean locations of the datapoints of each class and puts the decision boundary halfway between them



finds the decision boundary that separates two classes such that there is the SMALLEST space (margin) between the boundary and the nearest training samples (support vectors)

Question 4**0.1 / 0.1 pts**

Nonparametric statistics...



do not make assumptions about the underlying distribution generating the data.



are more sensitive than parametric (classical) statistics.



require less data than parametric (classical) statistics.



can only be used with Normally-distributed data.

Question 5**0.1 / 0.1 pts**

Nonparametric statistics are often helpful when analyzing...



continuous data



quantitative data



ordinal data



Normally-distributed data

Question 6**0.1 / 0.1 pts**

Which methods are typically more sensitive (able to correctly reject H_0 when the difference is small)?

- ☒ Parametric methods
- ☐ Nonparametric methods

Question 7**0.1 / 0.1 pts**

A neural net reflects a neuron in that...

- ☐ it uses distance to find the closest neighbour to predict the label for the given input.
- ☐ input to the decision tree branches that helps to grow the tree.
- ☒ it takes in bunch of inputs, processes it and determines the appropriate output.
- ☐ it is like the nerve that connects different part of the brain.

Question 8**0.1 / 0.1 pts**

Select all statements about neural networks that are true (select all)

☒ Can be used for supervised and unsupervised learning.



neurons in a network sum up their inputs and then apply a (usually) non-linear transformation to form the output.

☒ Can learn features useful for classifying images.

☐ Is the only effective way for classifying images.

Question 9

0.1 / 0.1 pts

After running a model on a test set, you generate a confusion matrix. How do you calculate the sensitivity of the model?

☐ True Negatives / (True Positives + False Negatives)

☐ True Negatives / (True Negatives + False Positives)

☒ True Positives / (False Negatives + True Positives)

☐ True Positives / (False Positives + True Positives)

Question 10

0.1 / 0.1 pts

If my overall model predictions are pretty good, a few extreme outliers will NOT affect my Root Mean Squared Error.

☐ True

☒ False

Quiz Score: **1** out of 1