Started on	Tuesday, August 13, 2019, 9:16 PM
State	Finished
Completed on	Tuesday, August 13, 2019, 9:18 PM
Time taken	1 min 38 secs
Points	17.00/17.00
Grade	**100.00** out of 100.00

Question **1**

Correct

1.00 points out of 1.00

By evaluating different ML algorithms run on the same data, you'll gain insights into the strengths and weaknesses of each one.

Select one:

- True ✔
- False

Yes, Lantz claims that at the top of p 312.

The correct answer is 'True'.

Question **2**

Correct

1.00 points out of 1.00

If we could ever get a classifier to have 99.99% accuracy, that would be great!

Select one:

- True
- False ✔

Not necessarily. As Lantz notes on p 312, that would be more true if the data itself isn't lopsided.

The correct answer is 'False'.

Question **3**

Correct

1.00 points out of 1.00

For best accuracy, usually we want to test with data in a live environment.

Select one:

- True
- False ✔

Maybe. Lantz describes on p 313 how usually this isn't feasible. We might want to, but we should have good alternatives available.

The correct answer is 'False'.

Question **4**

Correct

1.00 points out of 1.00

To evaluate a classifier, we should know its predictions, and also how certain it is about those predictions.

Select one:

- True ✔
- False

Yes, indeed. See p 314.

The correct answer is 'True'.

Question **5**

Correct

1.00 points out of 1.00

It's not important to dwell on mistakes an algorithms makes.

Select one:

- True
- False ✔

Actually, it is. See the examples on pp 316-317, of errors classifying "ham" and "spam."

The correct answer is 'False'.

Question **6**

Correct

1.00 points out of 1.00

The confusion matrix can be more than 2 by 2, and it is a basic tool for measuring an algorithm's performance.

Select one:

- True ✔
- False

Indeed. See p 317.

The correct answer is 'True'.

Question **7**

Correct

1.00 points out of 1.00

A false negative is incorrectly classified as the class of interest.

Select one:

- True
- False ✔

No, that's a false positive. See p 318.

The correct answer is 'False'.

Question **8**

Correct

1.00 points out of 1.00

The confusion matrix produced by the "caret" package adjusts accuracy by considering the possibility of a correct prediction by chance alone.

--> Homework: Take a snapshot of your confusion matrix for ham vs spam, like Lantz's on p 324, to turn in.

Select one:

- True ✔
- False

Yes, this is the "kappa" statistic described on p 323.

The correct answer is 'True'.

Question **9**

Correct

1.00 points out of 1.00

The goals of "sensitivity" vs "specificity" tend to be opposites -- it's hard to optimize both.

Select one:

- True ✔
- False

That's right -- as in the example on p 326, of getting rid of most of the "ham," just to be sure you got all the "spam."

The correct answer is 'True'.

Question **10**

Correct

1.00 points out of 1.00

You can even measure how interesting and relevant a model's results are!

Select one:

- True ✔
- False

Actually, you can. This is "precision" and "recall", introduced on p 328.

The correct answer is 'True'.

Question **11**

Correct

1.00 points out of 1.00

And, you can combine precision and recall into a single number!

Select one:

- True ✔
- False

Well, I suppose. That's what the F-measure claims to do. See p 330.

The correct answer is 'True'.

Question **12**

Correct

1.00 points out of 1.00

ROC curves show graphically the trade-off between false negatives and false positives.

Select one:

- True
- False ✔

I apologize for throwing another false one at you, about false negatives and false positives. They are important to distinguish. See p 332.

The correct answer is 'False'.

Question **13**

Correct

1.00 points out of 1.00

A ROC curve compares your classifier's performance versus both a perfect classifier and a useless classifier.

--> **Homework:** Take a snapshot of your ROC curve, like Lantz's on p 334, to turn in.

Select one:

- True ✔
- False

Yes, see p 332.

The correct answer is 'True'.

Question **14**

Correct

1.00 points out of 1.00

Measuring "resubstitution error" is done to see how far off an algorithm is, even on the data used for training.

Select one:

- True ✔
- False

Yes, that's it, but it's even better to guess at how far off the algorithm will be on data it's never seen. See p 336.

The correct answer is 'True'.

Question **15**

Correct

1.00 points out of 1.00

Having an unseen data set available is a luxury that's usually not available.

Select one:

- True
- False ✔

Yes, it is. You withhold a "validation dataset" from all the training done on the model, then use it as the unseen data. This is what we've been doing in most of the Modules of this course. See p 337.

The correct answer is 'False'.

Question **16**

Correct

1.00 points out of 1.00

In "k-fold cross-validation," you divide the data into completely separate random partitions, like 10 of these, say, train and test separately on each of them. This gives an "average" performance of the algorithm on this data.

Select one:

- True ✔
- False

Yes, this is how it's done. See p 340.

The correct answer is 'True'.

Question **17**

Correct

1.00 points out of 1.00

Bootstrapping means you start over, multiple times, with randomly picked training and test datasets.

Select one:

- True ✔
- False

Yes, indeed. See p 343.

The correct answer is 'True'.

Question **18**

Complete

Not graded

We want to base online and remote face-to-face discussions on the topics of most value to you.

Please think carefully about all the material you read, then write a prompt for discussion you would like to hear - either:

a. Something that you aren't sure about, which you'd like to have explained in class, or

b. A topic you liked a lot, that you'd like to discuss in class.

na

Thanks!