

Started on	Wednesday, June 19, 2019, 8:23 AM
State	Finished
Completed on	Wednesday, June 19, 2019, 9:14 AM
Time taken	51 mins 26 secs
Points	6.00/6.00
Grade	100.00 out of 100.00

Information

Lantz introduces the example of breast cancer diagnosis by discussing why this is important. The fact is, ML algorithms now play a major role in medical diagnoses. One study famously showed that, relying on these automated guesses, doctors actually did better than if they added their own guesses to that ML output!

You should pay close attention to the 5-Step model Lantz shows, in the headings for this second half of Ch 3. He'll use these same headings over and over. And, they represent the steps you need to go through in using most ML algorithms. Read this whole section, pp 75-87, following along by duplicating the coding yourself. Then take the quiz questions, below, to test your understanding.

Just a few tips as you read and try things:

- Bring up the Ch 3 data file in Excel (or the equivalent), so you can see what he's talking about, in "Step 1 -- collecting data", on p 76.
- The "features" Lantz is discussing are either a column in the spreadsheet, or, sometimes, a combination of those.
- If your RStudio is pointing to the Ch3 directory as the working directory, the "read.csv()" on p 77 ought to work!
- Notice the data frame operations Lantz uses to edit the data he read-in. Like how to "drop a feature," at the bottom of p 77. Or, how to "recode" features as "factors" on p 78. Or, how to change the scale of a variable with "round()", on p 78.
- On p 79, you get to try some real programming! You define the "normalize()" function yourself, and then, as Lantz says, you can use it anywhere. I hope you can imagine what a general feature this is. You could define any function you like, basically, and then use it. Of course, there are rules, like the use of a "return" in the function definition, to say what value it gives back.
- The last line on p 79 might look like magic, but Lantz explains it at the top of the next page. Please think this over and, if you don't feel you get it completely, start a topic about it on the Module 3 Forum. Call it "p 79 magic".
- Notice how Lantz verifies every major data transformation, to verify it worked correctly. This is great QA! Like, on p 85, after he transforms the data to z-scale standardization, he checks the results to see how close the mean is to zero.

--> To save for homework: Copy your CrossTable results from p 83, and also from p 86, into a file, to turn in as Module 3 Homework.

Question **1**
Correct
1.00 points out of 1.00

In general, the more "features" you include in the k-NN training, the better.

Select one:

- ☐ True
- ☒ False ✓

No, on p 76 Lantz warns about including features you don't know are related to the classification you are doing, and on p 77 he warns about including ID variables.

The correct answer is 'False'.

Question **2**

Correct

1.00 points out of 1.00

The reason we divide the data into "training" and "test" datasets is because otherwise we are just predicting what we already know.

Select one:

- ☒ True ✓
- ☐ False

That's how Lantz starts his discussion, middle of p 80.

The correct answer is 'True'.

Question **3**

Correct

1.00 points out of 1.00

In normalizing the data frame, we excluded column 1 because it was the ID.

Select one:

- ☐ True
- ☒ False ✓

No, this is the diagnosis, which is brought into a "labels" data frame on p 81. We had already gotten rid of the ID column, bottom of p 77.

The correct answer is 'False'.

Question **4**

Correct

1.00 points out of 1.00

Running the "knn()" classifier function, we end up with a vector of factors predicting the outcomes for each row of the test data frame.

Select one:

- ☒ True ✓
- ☐ False

Yes, that's what it does. p 83.

The correct answer is 'True'.

Question **5**

Correct

1.00 points out of 1.00

The results of the k-NN run, on p 83, are nearly ideal, because there were no cases predicted to be Malignant which were actually Benign.

Select one:

- ☐ True
- ☒ False ✓

No, that's good that we got this prediction. However, we also predicted two "false negatives" -- in the lower left quadrant of the table. These predicted Benign when they were actually Malignant.

The correct answer is 'False'.

Question **6**

Correct

1.00 points out
of 1.00

Lantz's Step 5 -- Improving model performance was very successful. The results were more accurate.

Select one:

- ☐ True
- ☒ False ✓

On the contrary, there were 5 false negatives, instead of 2!

The correct answer is 'False'.

Question **7**

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.

N/A

Thanks!