Started on	Wednesday, June 19, 2019, 7:56 AM
State	Finished
Completed on	Wednesday, June 19, 2019, 8:10 AM
Time taken	13 mins 30 secs
Points	7.00/7.00
Grade	**100.00** out of 100.00

Information

This is exciting stuff -- you get to see the general theme for all the succeeding chapters in Lantz's book!

And, you can begin to understand the general approach for using any ML algorithm on your data!

And, all of this is while you learn your first, fairly straightforward algorithm -- "Nearest Neighbor."

To begin each chapter, Lantz describes what's going on. Then he has a detailed example, where you get to try programming the algorithm on the data that he has provided.

Start by reading the descriptive stuff. That's pp 65-75. See if you can go over it till you have a basic understanding, then try the first group of quiz questions, below.

Question **1**

Correct

1.00 points out of 1.00

Lantz says "nearest neighbor" is a classifier algorithm where "you'll know it when you see it," because it can take messy data and divide it into homogeneous classes.

Select one:

- True ✔
- False

That's exactly what he's saying, in the middle of p 66.

The correct answer is 'True'.

Question **2**

Correct

1.00 points out of 1.00

Looking at the list o strengths and weaknesses of k-NN, on p 67, you could describe it as "quick and dirty."

Select one:

- True ✔
- False

That's fair. It is easy to use, makes no assumptions, and has fast "training." (That's how quickly it learns what to do.)

The correct answer is 'True'.

Question **3**

Correct

1.00 points out of 1.00

In Lantz's food classifier system, the 1-NN algorithm called tomatoes a fruit, because they were closest to an orange, in sweetness and crunchiness, using Euclidean distance.

Select one:

- True ✔
- False

That's exactly how he decide it, on p 70.

The correct answer is 'True'.

Question **4**

Correct

1.00 points out of 1.00

In the k-NN algorithm, picking k to be the entire number of observations, in the training data, is a good idea because it reduces the impact or variance caused by noisy data.

Select one:

- True
- False ✔

No, that's too big. As Lantz says on p 71, this would cause the algorithm always to predict that any new observations were in the majority class.

The correct answer is 'False'.

Question **5**

Correct

1.00 points out of 1.00

It's likely your data will have to be "prepared" by adjusting the scales, for different variables, to be close to the same.

Select one:

- True ✔
- False

Yes, otherwise, as Lantz notes on p 72, "spiciness rules!"

The correct answer is 'True'.

Question **6**

Correct

1.00 points out of 1.00

"Nominal data" is data like male vs female, which is not numeric. We can handle that, by using "dummy coding."

Select one:

- True ✔
- False

Yes, that's the trick described on p 73. Make one of them = 1, and the other = 0, say.

The correct answer is 'True'.

Question **7**

Correct

1.00 points out of 1.00

The k-NN algorithm is lazy because it doesn't create any abstraction of what's going on -- you could say it "doesn't learn anything."

Select one:

- True ✔
- False

That sums-up what Lantz says, on pp 74-75.

The correct answer is 'True'.

Question **8**

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.

Is there a better way to classify Nominal data. Dummy coding seems bit inefficient.

Thanks!