Riley Payung

CDS 303

03/16/2020

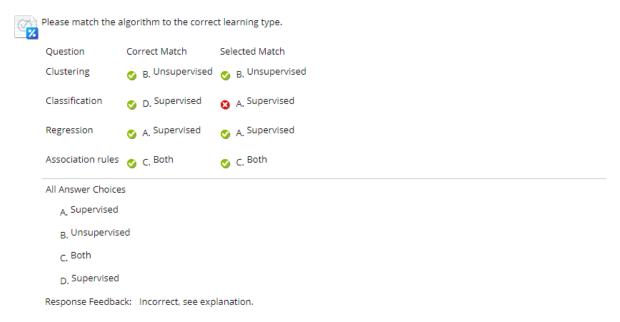
Redos

Quiz 2

Q2.

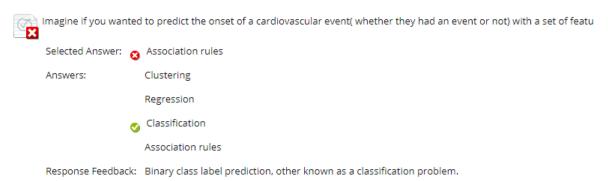
I did not actually get anything wrong on this question, just chose the wrong option, where two options were the same.

Question 2



Q3.

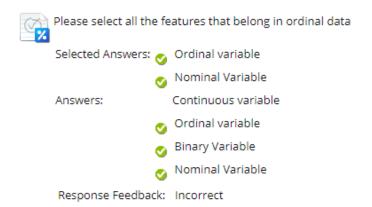
Simple Mistake, I chose Classification when I should have chosen Association Rules.



Q4.

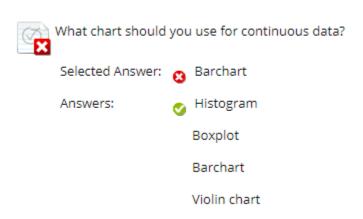
I chose only Ordinal and Nominal, when I should have chosen Ordinal, Nominal, AND binary.

Question 4



Q6.

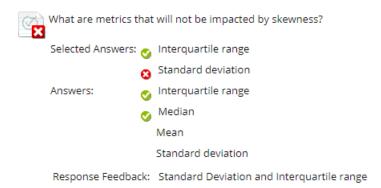
I chose Bar chart instead of Histogram. I think that I was frantically trying to get this quiz done because that was just a dumb mistake on my part.



Q7.

Again, just a dumb mistake that I should not have made. I chose Standard Deviation when I should have chosen median.

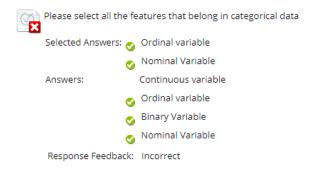
Question 7



Quiz 3

Q1.

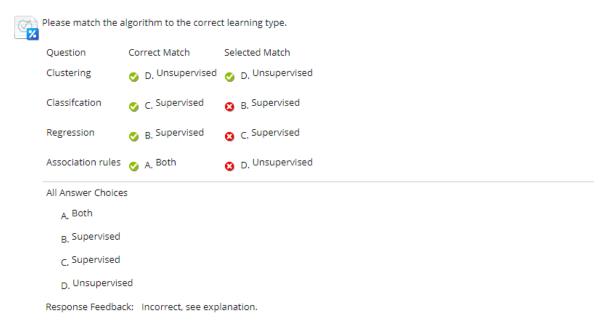
I did the same thing I did in Quiz 2. I should have also chosen Binary.



Q2.

Same as Quiz 2 except for Association Rules, I should have chosen both. I believe that I had issues with time management on this one.

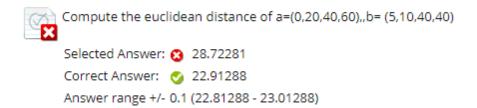
Question 2



For Questions 5 through 10, I did not do well on them because I took the quiz without going through the notes, and later when I was studying for the midterm, I found what I needed to complete these questions in the notes. Until that point, I was completely ignorant.

Q5.

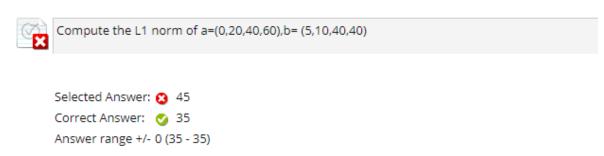
I somehow messed up Euclidean Distance, which is the generic distance formula we use. I don't remember how I calculated it or how I managed to get it wrong, but I got 28.72281 when I should have got 22.912.



Q6.

This one was kind of interesting because I've never seen the Manhattan Distance (L_1 Norm) before. I should have been able to figure this one out though, but again, not really sure what I was thinking at the time of taking this quiz.

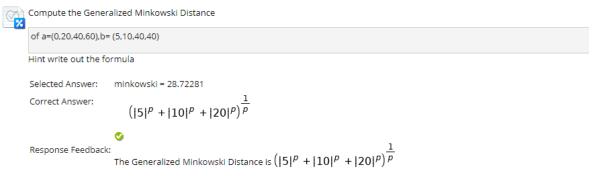
Question 6



Q8.

I should have wrote out the formula but I did not; When I was answering this question, I think that I was confused as to what to use for p, but seeing the answer made me realize that I should have just wrote out the formula.

Question 8



Q10.

Again, I did not pay enough attention to the notes. I did not understand the idea of proximity measures as well as I thought I did.



Quiz 4

01

I chose the wrong formula, when I should have chosen Circumference~age. The new model does in fact fit better.

```
call:
lm(formula = Orange$circumference ~ Orange$age, data = Orange)
Residuals:
                           1Q
                                   Median
        Min
                                                                            Max
-46.310 -14.946
                                   -0.076 19.697
Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept) 17.399650
                                                   8.622660 2.018 0.0518
                                                   0.008277 12.900 1.93e-14 ***
                        0.106770
Orange$age
Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
Residual standard error: 23.74 on 33 degrees of freedom
Multiple R-squared: 0.8345, Adjusted R-squared: 0.8 F-statistic: 166.4 on 1 and 33 DF, p-value: 1.931e-14
                                                              Adjusted R-squared: 0.8295
  Ouestion 1
           Using the Im, and Summary function in R and the Orange data set, I would like you to answer the following. If the outcome is circumference and the feature is age. Please report the slope, intercept.
      Using the initiation of the slope.

Please tell me the interpretation of the slope.
           Please tell me the amount of variance explained by the model.
           If you were to fit a new model with Tree and age as features, does your model fit any better? What is your justification for whether or not it fits better?
                            Age - Min: 118.0, 1st O.: 484.0, Median: 1004.0, Mean: 922.1, 3rd O.: 1372.0, Max: 1582.0
                           Age - Mill: 118.0, 151 Q2. 494-0, Media in: 1004-0, Media: 922-1, 310 Q2. 1572-0, Max. 1302-0
Circumference - Miln: 3.0., 151 Q2. 65.5, Media: 115.0, Mean: 115.9, 310 Q2. 161.5, Max: 214.0
The intercept is 16.604 and the slope is 7.816. Since the correlation of the model is 0.9135, it is a strongly positive model with low variance.
                           The model would not fit any better. The intercept is 9.221e+02 (922.1) and there are multiple slopes.
           Correct Answer:
                              Intercept 17.399650
                              Slope 0.106770
                              For every year a tree ages the circumferences increases by 0.11 cm
                              The variance explained by models is the R^2 0.8345. So 83% of the variance is explained by the model > summary(Im(circumference-age,df1))
                              Residual standard error: 23.74 on 33 degrees of freedomMultiple R-squared: , Adjusted R-squared: 0.8295 F-statistic: 166.4 on 1 and 33 DF, p-value: 1.931e-14
                           This will give you an intercept of 17.40 and slope of 0.11. This tells you that for every year a tree ages the circumferences increases by 0.11 cm.
                           Also, the variance explained by models is the R^2 0.8345. So, 83% of the variance is explained by the model.
```

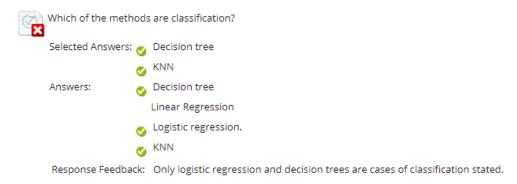
The new model is a better fit because the amount of variance explained by the model increases.

Quiz 5

Q2.

This is another question in which I should have followed the notes a little more carefully, but I did not choose one of the options that I should have. I should have also chosen Logistic Regression, but I did not.

Question 2



Lab #3

I did not do the lab. I should have done the lab, because now that I am doing it, it seems quite easy to do, and I missed out on a whole extra 5 points that I could have received from finishing it on time. I think this was a week that I had very poor time management.

I have submitted the lab under the original assignment and will upload the same pdf along with this document.

Week #3 Discussion Article

I did not do the discussion article because I was attempting to mitigate damage to my grade.

Week #6 Discussion Article

I did not do the discussion article because the discussion board was not available at the time. I should have emailed Professor Valko about this when I saw it, but I did not.