← BACK TO GRADES

Homework Assignment 6 (week 7)

Faculty Feedback

Score Last published: 8/29/2019 10:53 AM PDT

100 / 100 (100.00%)

Comments

Introduction:

WOW! Great intro. You have really set the stage and have clearly identified the problem of handwriting recognition and have provided a clear segue to the investigated method – Decision Trees + Bayes.

EDA + Analysis:

The overall discussion of the **handwriting** data, cleaning and eda was great!

Results:

WOW -- Great decision trees and Bayes Classifiers!!! The results comparison and visualizations are excellent! I

Great job comparing multiple parameterizations of the classifiers!!

I really like the way you presented the results. It made comparisons very easy to understand.

Conclusions:

Overall a good submission -- nice job!

Feel free to review the following to think about areas that can use improvement or areas that might also need improvement when you create your project write-up.

Remember – the project write up is very similar to the Assignments – but with a lot more writing, models, methods, measures, results, etc.

As we have moved through this class, **your work has been both excellent and has shown a clearer and improved understanding** of data science and of related concepts.

The following will review the general headings and requirements for this assignment.

My grading is holistic and I encourage you to review all of the following and always ask yourself how you can continue to improve.

General notes to assist with self-designated improvement:

Required Headings....

Introduction: (2 - 3 paragraphs)

An Introduction is about the area or topic, not about the data or models. The introduction helps the reader to understand what the assignment area is about. For example, suppose the assignment is about schools. In this case, the introduction is about school systems, why schools are measured and ranked, who might be concerned with school measures and rankings (such as students, parents, states, governments, and funding agencies), and the value of comparing schools.

An introduction is a like a warm-up or like dating. It allows the reader to “get to know” the area of interest.

Extra Notes

(1) An Intro should be a comparative size with respect to the paper. For an Assignment, a good Intro is 3 paragraphs. For a project, it will be 5 – 6.

(2) Writing skills are never easy and all people can improve. The Introduction not only introduces your topic to the reader, but it also introduces your credibility. Can they believe you? If your writing is very good, this can help. If your writing is not as good, this can hinder.

(3) If an Introduction has a “undergrad or studenty” feel, such as starting sentences with “My Assignment is about…”, or “I am going to talk about…”, or “In this Assignment, I will….”, this will make you appear to be a student rather than an expert or professional.

- At the graduate level, write Assignments as though they are technical papers. This means the paper is never about you or your experience, but rather about the topic (such as pollution, mushrooms, dolphins, or diabetes).

- Avoid saying things like, “I tried to use technique X and had this experience….” Rather, show all models and methods (and measures) and the Results section, discuss each.

(4) Avoid the use of “I”, “we”, “us”, “you”, … Always remember that Assignments are about the area of interest, the topics, the models, the analysis, the results, and the outcomes/conclusions, they are not about you.

(5) The Intro should not contain any information about the dataset or the data cleaning, prep, processing, etc. Everything about the dataset goes into the Analysis section under the “About the Data” subsection.

Analysis and Models

The Analysis section contains subsections.

The first subsection is “**About the Data”** which contains all the information about the dataset, the variables, the cleaning and prep, checking for and dealing with missing values, checking for and dealing with incorrect values, checking for an dealing with outliers, feature generation, normalization (if needed), etc.

In this subsection, you will also “explore” the data. This means that you write about each variable, visualize each variable, and talk about what the variable represents. Tables are great for this as well.

Data formatting, cleaning, preparation, feature generation, normalization, etc. should be clear, transparent, explained, and visualized (as possible).

The second and remaining subsections of Analysis are the model(s).

In some cases, there may only be one model.

A model is any method used to analyze the data.

Each Assignment specifies which models to use.

Always include model details and parameter values when applicable.

*** Have Visualizations throughout the assignment.

Results

The Results section of the Assignment will have a subsection for results for each model (assuming that you have more than one).

Results are technical.

They offer technical information about what was found in the analysis.

For example, if you performed a correlation in the analysis between all pairs of numeric variables, then your results would discuss the r-value and relationship of each pair. Similarly, if you looked at measures of center and variation, the results talk about what those measures are and what they reveal. For example, if the mean is less than the median, the data is skewed, which means....

Each model we will use in this class has results and parameters associated with it. For example, association rule mining will offer the top ten rules for sup, conf, and/or lift if you you code it to do so. These would go into the results along with the sup, conf, and lift for each rule. The meaning would also be discussed.

** Always have visualizations

Conclusions

2 - 3 paragraphs (at least)

This area is not technical at all.

For example, the Conclusions should never contain words like Association Rules, Confidence, k – means, trees, kernels, etc. etc.

This area explains what was actually found - in direct relation to the topic – and in a way that would make sense to anyone.

For example, if you discovered in the analysis that association rule mining with a conf of .2 and a sup of .3 offered 10 rules, you would talk about the measures and values and rules in

the *Results*. In the Conclusions, you would talk about what it all means. So you would not include the rules themselves or mention of technical measures such as conf or sup. Rather, you would say that you found (as a random example) that people who buy diapers are very likely to buy beer and that this means that a store should consider placing these items “near” to each other.

Basis For Grades:

100: This means that your Assignment was amazing and so perfect that nothing can be improved. It covered everything – cleaning, prep – analysis that makes sense – visualizations – results (that are true) – etc. There is nothing really left to improve.

95: This means that your Assignment is really good! You covered most of the items noted below and perhaps a few others not noted. You can make some improvement on pre-processing and results analysis, as well as perhaps other visualizations. Overall – you have the idea and you did well.

90: This means that your Assignment is good, but could be a little better. Perhaps add items such as further data cleaning and pre-processing, data normalization, better or more visualizations, and/or more robust conclusions. Many students forgot to change Section to a factor for example. Very few students summed and normalized the data to look at the percentages for each attribute.

85: This means that your Assignment is a good start and largely meets the more general and overall requirements. Here, you used R, you did some analysis, you did some cleaning, you made some graphs, and your reached some conclusions. However, there is room for improvement.

Below 85 means that the level of 85 above was not quite met and many elements were missing.

Extra Assignment – Specific Notes:

DT

NB

Measures and comparisons

Performance

Are the models constructed correctly?

Is the result analysis conclusion convincing?

Is sufficient details provided for others to repeat the analysis?

Does the analysis include irrelevant content?

Successful submission to Kaggle? - optional

Student Submission | Homework Assignment 6 (week 7)

Response

Last submitted: 8/21/2019 2:27 PM PDT

No response

Files | Download all (4)

File Name	Uploaded	Feedback
Ryan_Timbrook_HW6.docx	8/21/2019 2:23 PM PDT	
ist707_w7_hw6.ipynb	8/21/2019 2:23 PM PDT	
submission.csv	8/21/2019 2:24 PM PDT	
ist707_w7_hw6.zip	8/21/2019 2:27 PM PDT	