CS688 Assignment 4: Classifier Metrics, Shiny

Please follow the submission requirements at the end of the assignment!

Objectives: First, be able to calculate a confusion matrix and understand other metrics for classifier effectiveness. Second, demonstrate your ability to create a Shiny app.

Part A (25 points)

1. (15 points) For the newsgroup classification problem in Assignment 3, estimate the effectiveness of your classification:
   - Create the confusion matrix (identify which value is TP, which is TN, which is FP, which is FN)
   - Calculate Precision
   - Calculate Recall
   - Calculate F-score
2. (10 points) Discuss whether the values you got for TP, TN, FP, FN, precision, recall, and F-score are "good".

Part B (75 points)

Modify the Shiny Wikipedia Search App shown in example 4.3 to:

1. (45 points) Replace the current visual display (dendrogram) with a wordcloud() of the 50 most frequent words found in the articles retrieved from Wikipedia.
   - It might help to decide what parts of server.R and/or wikiSearch.R create a dendogram (and delete these parts), while keeping the parts of server.R and/or wikiSearch.R that will help you create a wordcloud.
   - Some of you might prefer to combine wikiSearch.R and server.R into a single file, but this is not required.
2. (30 points) Add functionality to the same app to display the 50 most frequent terms and the number of times they occur in the Wikipedia web pages that were retrieved based on the user selection. You need to put the words into a data frame and learn how to output a data frame in Shiny. Consider the function tableOutput() and renderTable().
   - Ensure the same words show in the wordcloud and the table.

Part C (5 points extra credit)

In past sessions, many students have attempted to solve part B, problem 1 with code that essentially simplifies to plot(wordcloud(...)). Explain why this does not work.

SUBMISSION REQUIREMENTS:

- **Create a Word, PDF, or Rmd document. If you use Rmd you will need to make sure to save the output as a PDF.**
- **For each question, state the question you are answering. Then answer the question by explaining in sentences (in English, not in R or other languages) what you did to get to the answer. You may include screenshots and/or copy-paste of key lines of code and the**

corresponding output in your answer. (If you are using Rmd, this means you must generally use echo=FALSE and/or include=FALSE for the body of the document.)

- Full code should be included as an Appendix to your Word or PDF document. Coding must be in R. Do NOT include full code in the main part of your document.
- Please ensure that a Word or PDF file as the first file in your submission.
- You may also separately upload your R and/or Rmd code to Blackboard.
- If your facilitator tells you to submit the files differently than the above guidelines, you are expected to respect your facilitator's wishes starting on the next assignment.
- Facilitators can deduct up to 20% if you fail to follow these requirements (more if the questions are not actually answered).
- Facilitators can deduct 5% for each day the assignment is late. You may submit one (and only one) of the six assignments up to three days late with no penalty but all other assignments will be penalized.
- Unless your facilitator or the professor agrees, your assignment will not be graded if it is more than 3 days late (e.g., no credit will be given after Friday at 6 AM Boston time). The professor will usually ask the facilitator to make the decision but in rare cases (<1% of the time) has overridden a facilitator. Do not expect the professor to override in most cases.

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder