Assignment

Text Classification Problem

Steps

1. You have to identify a text classification problem and dataset from UCI text datasets

You are responsible for splitting the data into two parts: train and test

2. You have to preprocess the data. It involves removing stop words, converting words to vectors, choosing how many words you would like to use and finally make the data suitable for being used as an input layer.

Hint: It's totally upto you how you do this. A good resource is the *tm package* in R

Text Classification Problem

Steps

3. Create a deep network using Keras package. It's upto you to design this. Remember to design the output layer based on the data features.
4. **You need to run your code on Google Cloud with GPU enabled. We had given you an introduction in class. You will need to figure out the details - such as how to transfer to cloud and refer to them in your code**

Some useful sites:

https://cloud.google.com/storage/docs/creating-buckets

gs_copy function of cloudml package
https://cran.r-project.org/web/packages/cloudml/cloudml.pdf

Text Classification Problem

Steps

5. Split the **train** data into train and validation parts. Use the fit function to build and validate your data and create a plot of the history.
6. Apply the model on the test part and see how well it does.
7. How can you improve the performance. You need to vary the parameters and run your code at least 5 times and report the parameters used and accuracy obtained.