BigML Assignment 4a: Naive Bayes using Hadoop Streaming

Runyun Zhang

runyunz@andrew.cmu.edu

1. **Examine the output of your Hadoop streaming job on the full RCV1 training dataset. What are the counts for the following keys? (20 points)**

   Notice: case-sensitive

   Y=GCAT,W=gunman 9

   Y=GCAT,W=half 17109

   Y=CCAT,W=limited 47

2. **Your answers to collaboration policy (on the first page of this handout).**

   Did you receive any help whatsoever from anyone in solving this assignment? No.

   Did you give any help whatsoever to anyone in solving this assignment? No.