Running MLLIB EXAMPLE CODE

Scala already comes with a large collection of Machine Learning code examples. They are a great source for learning Scala based programming as well as ML techniques.

Location of examples:
/usr/local/spark-1.4.1-bin-hadoop2.4/examples/src/main/scala/org/apache/spark/examples

How to run examples:
- You can run all of the examples from the command line as follows:

      run-example <ClassName> <Parameters>

For example,

      run-example SparkPi 10

Another example,

      run-example HdfsTest <any file on HDFS>

- The tricky part might be understanding all the parameters needed for the program

Running MLLIB EXAMPLE CODE

To run any of the code examples from the MLlib examples:
http://spark.apache.org/docs/latest/ml-guide.html

you need to follow the steps below:

1. Copy the file to cs6360 cluster.
2. If you would like to use a data file from the examples, make the following change:

Replace **data/** with
file:////usr/local/spark-1.6.1-bin-hadoop2.4/data/

3. Replace **sqlCtx** with **sqlContext**

4. Add the following at the end of the code:
exit

5. If you don't want to see a lot of log messages, type the following at the start of the code:
sc.setLogLevel("WARN")