LOGS FOR ALL 4 FILES, sampler.py, trainer.py, tester.py and log\_extension.py

The errors in between are due to repeated commands execution, I did not overwrite anything to be at a safer side.

[ag6516@login-1-1 anshul\_project]$ spark-submit --driver-memory 8g --executor-memory 8g sampler.py

20/01/30 00:52:46 WARN Client: Neither spark.yarn.jars nor spark.yarn.archive is set, falling back to uploading libraries under SPARK\_HOME.

Traceback (most recent call last):

File "/home/ag6516/anshul\_project/sampler.py", line 57, in <module>

main(spark)

File "/home/ag6516/anshul\_project/sampler.py", line 34, in main

model\_uid.save('anshul\_project/model\_uid')

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/pyspark.zip/pyspark/ml/util.py", line 244, in save

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/pyspark.zip/pyspark/ml/util.py", line 183, in save

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/py4j-0.10.7-src.zip/py4j/java\_gateway.py", line 1257, in \_\_call\_\_

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/pyspark.zip/pyspark/sql/utils.py", line 63, in deco

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/py4j-0.10.7-src.zip/py4j/protocol.py", line 328, in get\_return\_value

py4j.protocol.Py4JJavaError: An error occurred while calling o244.save.

: java.io.IOException: Path anshul\_project/model\_uid already exists. To overwrite it, please use write.overwrite().save(path) for Scala and use write().overwrite().save(path) for Java and Python.

at org.apache.spark.ml.util.FileSystemOverwrite.handleOverwrite(ReadWrite.scala:702)

at org.apache.spark.ml.util.MLWriter.save(ReadWrite.scala:179)

at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)

at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)

at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)

at java.lang.reflect.Method.invoke(Method.java:498)

at py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)

at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.java:357)

at py4j.Gateway.invoke(Gateway.java:282)

at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.java:132)

at py4j.commands.CallCommand.execute(CallCommand.java:79)

at py4j.GatewayConnection.run(GatewayConnection.java:238)

at java.lang.Thread.run(Thread.java:748)

[ag6516@login-1-1 anshul\_project]$ spark-submit --driver-memory 8g --executor-memory 8g trainer.py

20/01/30 00:56:34 WARN Client: Neither spark.yarn.jars nor spark.yarn.archive is set, falling back to uploading libraries under SPARK\_HOME.

20/01/30 01:01:38 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeSystemBLAS

20/01/30 01:01:38 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeRefBLAS

New best model

Rank: 4, Reg: 1, Alpha: 0.5

Traceback (most recent call last):

File "/home/ag6516/anshul\_project/trainer.py", line 52, in <module>

main(spark)

File "/home/ag6516/anshul\_project/trainer.py", line 45, in main

best\_model.save('anshul\_project/als\_sampling')

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/pyspark.zip/pyspark/ml/util.py", line 244, in save

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/pyspark.zip/pyspark/ml/util.py", line 183, in save

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/py4j-0.10.7-src.zip/py4j/java\_gateway.py", line 1257, in \_\_call\_\_

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/pyspark.zip/pyspark/sql/utils.py", line 63, in deco

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/py4j-0.10.7-src.zip/py4j/protocol.py", line 328, in get\_return\_value

py4j.protocol.Py4JJavaError: An error occurred while calling o185.save.

: java.io.IOException: Path anshul\_project/als\_sampling already exists. To overwrite it, please use write.overwrite().save(path) for Scala and use write().overwrite().save(path) for Java and Python.

at org.apache.spark.ml.util.FileSystemOverwrite.handleOverwrite(ReadWrite.scala:702)

at org.apache.spark.ml.util.MLWriter.save(ReadWrite.scala:179)

at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)

at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)

at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)

at java.lang.reflect.Method.invoke(Method.java:498)

at py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)

at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.java:357)

at py4j.Gateway.invoke(Gateway.java:282)

at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.java:132)

at py4j.commands.CallCommand.execute(CallCommand.java:79)

at py4j.GatewayConnection.run(GatewayConnection.java:238)

at java.lang.Thread.run(Thread.java:748)

[ag6516@login-1-1 anshul\_project]$ spark-submit --driver-memory 8g --executor-memory 8g tester.py

20/01/30 01:07:27 WARN Client: Neither spark.yarn.jars nor spark.yarn.archive is set, falling back to uploading libraries under SPARK\_HOME.

Test rmse 8.159281940361623

[ag6516@login-1-1 anshul\_project]$ spark-submit --driver-memory 8g --executor-memory 8g log\_extension.py

20/01/30 01:24:10 WARN Client: Neither spark.yarn.jars nor spark.yarn.archive is set, falling back to uploading libraries under SPARK\_HOME.

model created

pipeline created

20/01/30 01:29:20 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeSystemBLAS

20/01/30 01:29:20 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeRefBLAS

model fit

predictions created

evaluator created

evaluation 7.71668081677277

The best hyper parameters: Rank: 4, Reg: 0.5, Alpha: 1, RMSE: 7.71668081677277

Traceback (most recent call last):

File "/home/ag6516/anshul\_project/log\_extension.py", line 63, in <module>

main(spark)

File "/home/ag6516/anshul\_project/log\_extension.py", line 57, in main

als\_model.save('anshul\_project/log\_model')

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/pyspark.zip/pyspark/ml/util.py", line 244, in save

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/pyspark.zip/pyspark/ml/util.py", line 183, in save

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/py4j-0.10.7-src.zip/py4j/java\_gateway.py", line 1257, in \_\_call\_\_

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/pyspark.zip/pyspark/sql/utils.py", line 63, in deco

File "/share/apps/spark/spark-2.4.0-bin-hadoop2.6/python/lib/py4j-0.10.7-src.zip/py4j/protocol.py", line 328, in get\_return\_value

py4j.protocol.Py4JJavaError: An error occurred while calling o352.save.

: java.io.IOException: Path anshul\_project/log\_model already exists. To overwrite it, please use write.overwrite().save(path) for Scala and use write().overwrite().save(path) for Java and Python.

at org.apache.spark.ml.util.FileSystemOverwrite.handleOverwrite(ReadWrite.scala:702)

at org.apache.spark.ml.util.MLWriter.save(ReadWrite.scala:179)

at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)

at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)

at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)

at java.lang.reflect.Method.invoke(Method.java:498)

at py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)

at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.java:357)

at py4j.Gateway.invoke(Gateway.java:282)

at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.java:132)

at py4j.commands.CallCommand.execute(CallCommand.java:79)

at py4j.GatewayConnection.run(GatewayConnection.java:238)

at java.lang.Thread.run(Thread.java:748)