1. **Test environment**

|  |  |
| --- | --- |
| Hardware Information | SG2042 |
| Architecture | RISC-V64 |
| operating system | openEuler 24.03 (LTS) |
| Python Version | 3.1 2 . 2 |
| Java Version | OpenJDK 17 |
| javac version | 17 |
| GCC Version | 12.3.1 |
| G++ Version | 12.3.1 |
| Spark Version | 3.5.1 |
| Maven Version | 3.8.8 |
| Kernel version | 6.6.0 |
| Perf Version | 6.6.0 |
| cmake | 3. 27 . 9 |

1. **Test dependency preparation**

Use case address: attached to the Spark source code

Download address:

The specific operations are as follows:

1. Download and install the dependencies required to run the Python test case:

|  |
| --- |
| wget <https://archive.apache.org/dist/spark/spark-3.5.1/spark-3.5.1.tgz>  tar -xvzf spark-3.5.1.tgz  # Activate the venv virtual environment  source ~/venv\_3.1 2 \_pyspark/bin/activate  # ARM architecture pip installation dependencies  pip install torch scikit-learn grpcio protobu f grpcio-status torcheval  # RISC-V architecture pip installation dependencies  pip install scikit-learn grpcio protobu f grpcio-status torcheval parameterized  #Note : You also need to install torch, but the riscv version can only be installed from source code  # Currently tested, version 2.3.0 can run, higher versions are not guaranteed  #For details, see: <https://pypi.org/project/pytorch-riscv64/>  # Or csdn blog: https://blog.csdn.net/m0\_49267873/article/details/135670989  # Note that it is best to enter the spark-3.5.1 folder to run the test case  cd spark-3.5.1  # Note 1: When testing, we choose to use a single thread and do not use any acceleration library, so we need to add two lines of code in the conf/spark-default.conf file  # First, create spark-default.conf  cd conf/  cp spark-defaults.conf.template spark-defaults.conf  nano spark-defaults.conf  # Add the following two instructions to the file to disable openBlas:  spark.driver.extraJavaOptions=-Ddev.ludovic.netlib.blas.nativeLib=javaBLAS -Ddev.ludvic.netlib.lapack.nativeLib=javaLAPACK -Ddev.ludovic.netlib.arpack.nativeLib=javaARPACK  spark.executor.extraJavaOptions=-Ddev.ludovic.netlib.blas.nativeLib=javaBLAS -Ddev.ludovic.netlib.lapack.nativelib=javaLAPACK -Ddev.ludovic.netlib.arpack.nativelib=javaARPACK  #Note 2: When running Java test cases, a lot of INFO information will be output, which needs to be turned off to improve test performance.  #First go to spark-3.5.1/conf and create 'log4j2.properties' file from 'log4j2.properties.template'  #Then find 'rootLogger.level = info' in the file and change it to rootLogger.level = error  cp ./conf/log4j2.properties.template ./conf/log4j2.properties​​​​​​  sed -i 's/rootLogger.level = info/rootLogger.level = error/g' log4j2.properties |

1. **Use case testing**
2. Python test case 1: test case in ml/folder

Run case 1: aft\_survival\_regression

|  |
| --- |
| python ./examples/src/main/python/ml/aft\_survival\_regression .py |

Operation results:

|  |
| --- |
| Coefficients: [-0.4963068060199851,0.19844393975928465]  Intercept: 2.6380905631560196  Scale: 1.5472326865488455  +-----+------+--------------+------------------+---------------------------------------+  |label|censor|features |prediction |quantiles |  +-----+------+--------------+------------------+---------------------------------------+  |1.218|1.0 |[1.56,-0.605] |5.718996553029897 |[1.1603295951029087,4.9954717337196435]|  |2.949|0.0 |[0.346,2.158] |18.076458028588913|[3.6675401061563893,15.789559285491208]|  |3.627|0.0 |[1.38,0.231] |7.381875365763504 |[1.4977117707333796,6.447975512763028] |  |0.273|1.0 |[0.52,1.151] |13.57758129907789 |[2.754761130759771,11.859846908963412] |  |4.199|0.0 |[0.795,-0.226]|9.013093216625716 |[1.8286702406091513,7.872823838856867] |  +-----+------+--------------+------------------+---------------------------------------+ |

Run use case 2: als\_example

|  |
| --- |
| python ./examples/src/main/python/ml/ als\_example .py |

Operation results:

|  |
| --- |
| +------+--------------------+  |userId| recommendations|  +------+--------------------+  | 26|[{30, 6.2311215},...|  | 19|[{51, 3.6500263},...|  | 29|[{46, 4.968718}, ...|  +------+--------------------+  +-------+--------------------+  |movieId| recommendations|  +-------+--------------------+  | 65|[{27, 5.3176517},...|  | 26|[{10, 3.0312405},...|  | 29|[{2, 6.037142}, {...|  +-------+--------------------+ |

Run use case 3: binarizer\_example

|  |
| --- |
| python ./examples/src/main/python/ml/ binarizer\_example .py |

Operation results:

|  |
| --- |
| Binarizer output with Threshold = 0.500000  +---+-------+-----------------+  | id|feature|binarized\_feature|  +---+-------+-----------------+  | 0| 0.1| 0.0|  | 1| 0.8| 1.0|  | 2| 0.2| 0.0|  +---+-------+-----------------+ |

Run use case 4: bisecting\_k\_means\_example

|  |
| --- |
| python ./examples/src/main/python/ml/bisecting\_k\_means\_example .py |

Operation results:

|  |
| --- |
| Silhouette with squared euclidean distance = 0.9997530305375207  Cluster Centers:  [0.1 0.1 0.1]  [9.1 9.1 9.1] |

Run Example 5: bucketed\_random\_projection\_lsh\_example

|  |
| --- |
| python ./examples/src/main/python/ml/ bucketed\_random\_projection\_lsh\_example .py |

Operation results:

|  |
| --- |
| The hashed dataset where hashed values are stored in the column 'hashes':  +---+-----------+--------------------+  | id| features| hashes|  +---+-----------+--------------------+  | 0| [1.0,1.0]|[[-1.0], [-1.0], ...|  | 1| [1.0,-1.0]|[[0.0], [0.0], [0...|  | 2|[-1.0,-1.0]|[[0.0], [0.0], [-...|  | 3| [-1.0,1.0]|[[-1.0], [-1.0], ...|  +---+-----------+--------------------+  Approximately joining dfA and dfB on Euclidean distance smaller than 1.5:  +---+---+-----------------+  |idA|idB|EuclideanDistance|  +---+---+-----------------+  | 2| 5| 1.0|  | 0| 6| 1.0|  | 1| 4| 1.0|  | 3| 5| 1.0|  | 0| 4| 1.0|  | 2| 7| 1.0|  | 1| 7| 1.0|  | 3| 6| 1.0|  +---+---+-----------------+  Approximately searching dfA for 2 nearest neighbors of the key:  +---+----------+--------------------+-------+  | id| features| hashes|distCol|  +---+----------+--------------------+-------+  | 0| [1.0,1.0]|[[-1.0], [-1.0], ...| 1.0|  | 1|[1.0,-1.0]|[[0.0], [0.0], [0...| 1.0|  +---+----------+--------------------+-------+ |

Run Example 6: bucketizer\_example

|  |
| --- |
| python ./examples/src/main/python/ml/ bucketizer\_example .py |

Operation results:

|  |
| --- |
| Bucketizer output with 4 buckets  +--------+----------------+  |features|bucketedFeatures|  +--------+----------------+  |-999.9| 0.0|  | -0.5| 1.0|  | -0.3| 1.0|  | 0.0| 2.0|  | 0.2| 2.0|  | 999.9| 3.0| |

Run Example 7 : chisq\_selector\_example.py

|  |
| --- |
| python ./examples/src/main/python/ml/chisq\_selector\_example .py |

Operation results:

|  |
| --- |
| ChiSqSelector output with top 1 features selected  +---+------------------+-------+----------------+  | id| features|clicked|selectedFeatures|  +---+------------------+-------+----------------+  | 7|[0.0,0.0,18.0,1.0]| 1.0| [18.0]|  | 8|[0.0,1.0,12.0,0.0]| 0.0| [12.0]|  | 9|[1.0,0.0,15.0,0.1]| 0.0| [15.0]|  +---+------------------+-------+----------------+ |

Run example 8: chi\_square\_test\_example

|  |
| --- |
| python ./examples/src/main/python/ml/chi\_square\_test\_example .py |

Operation results:

|  |
| --- |
| pValues: [0.6872892787909721,0.6822703303362126]  degreesOfFreedom: [2, 3]  statistics: [0.75,1.5] |

Run use case 9: correlation\_example

|  |
| --- |
| python ./examples/src/main/python/ml/correlation\_example .py |

Operation results:

|  |
| --- |
| Pearson correlation matrix:  DenseMatrix([[1. , 0.05564149, nan, 0.40047142],  [0.05564149, 1. , nan, 0.91359586],  [ nan, nan, 1. , nan],  [0.40047142, 0.91359586, nan, 1. ]])  Spearman correlation matrix:  DenseMatrix([[1. , 0.10540926, nan, 0.4 ],  [0.10540926, 1. , nan, 0.9486833 ],  [ nan, nan, 1. , nan],  [0.4 , 0.9486833 , nan, 1. ]]) |

Run Example 10: count\_vectorizer\_example

|  |
| --- |
| python ./examples/src/main/python/ml/aft\_survival\_regression .py |

Operation results:

|  |
| --- |
| +---+---------------+-------------------------+  |id |words |features |  +---+---------------+-------------------------+  |0 |[a, b, c] |(3,[0,1,2],[1.0,1.0,1.0])|  |1 |[a, b, b, c, a]|(3,[0,1,2],[2.0,2.0,1.0])|  +---+---------------+-------------------------+ |

Run case 11: cross\_validator

|  |
| --- |
| python ./examples/src/main/python/ml/aft\_survival\_regression .py |

Operation results:

|  |
| --- |
| Row(id=4, text='spark ij k', probability=DenseVector([0.2665, 0.7335]), prediction=1.0)  Row(id=5, text='lm n', probability=DenseVector([0.9204, 0.0796]), prediction=0.0)  Row(id=6, text='mapreduce spark', probability=DenseVector([0.4438, 0.5562]), prediction=1.0)  Row(id=7, text='apache hadoop', probability=DenseVector([0.8587, 0.1413]), prediction=0.0) |

Run Example 12: dataframe\_example

|  |
| --- |
| python ./examples/src/main/python/ml/dataframe\_example .py |

Operation results:

|  |
| --- |
| Loading LIBSVM file with UDT from data/mllib/sample\_libsvm\_data.txt.  Schema from LIBSVM:  root  |-- label: double (nullable = true)  |-- features: vector (nullable = true)  Loaded training data as a DataFrame with 100 records.  +-------+-------------------+  |summary| label|  +-------+-------------------+  | count| 100|  |mean| 0.57|  |stddev|0.49756985195624287|  | min| 0.0|  | max| 1.0| |

Run Example 13 : dct\_example.py

|  |
| --- |
| python ./examples/src/main/python/ml/dct\_example .py |

Operation results:

|  |
| --- |
| +----------------------------------------------------------------+  |featuresDCT |  +----------------------------------------------------------------+  |[1.0,-1.1480502970952693,2.0000000000000004,-2.7716385975338604]|  |[-1.0,3.378492794482933,-7.000000000000001,2.9301512653149677] |  |[4.0,9.304453421915744,11.000000000000002,1.5579302036357163] |  +----------------------------------------------------------------+ |

Run use case 14 : decision\_tree\_classification\_example.py

|  |
| --- |
| python ./examples/src/main/python/ml/decision\_tree\_classification\_example .py |

Operation results:

|  |
| --- |
| +----------+------------+--------------------+  |prediction|indexedLabel| features|  +----------+------------+--------------------+  | 1.0| 1.0|(692,[121,122,123...|  | 1.0| 1.0|(692,[122,123,124...|  | 1.0| 1.0|(692,[124,125,126...|  | 1.0| 1.0|(692,[124,125,126...|  | 1.0| 1.0|(692,[124,125,126...|  +----------+------------+--------------------+  only showing top 5 rows  Test Error = 0  DecisionTreeClassificationModel: uid=DecisionTreeClassifier\_f0eb29433a1d, depth=2, numNodes=5, numClasses=2, numFeatures=692 |

Run use case 15: decision\_tree\_regression\_example

|  |
| --- |
| python ./examples/src/main/python/ml/decision\_tree\_regression\_example .py |

Operation results:

|  |
| --- |
| +----------+-----+--------------------+  |prediction|label| features|  +----------+-----+--------------------+  | 0.0| 0.0|(692,[98,99,100,1...|  | 0.0| 0.0|(692,[122,123,124...|  | 0.0| 0.0|(692,[122,123,148...|  | 0.0| 0.0|(692,[124,125,126...|  | 0.0| 0.0|(692,[124,125,126...|  +----------+-----+--------------------+  only showing top 5 rows  Root Mean Squared Error (RMSE) on test data = 0.149071  DecisionTreeRegressionModel: uid=DecisionTreeRegressor\_a56d0a53d1a1, depth=1, numNodes=3, numFeatures=692 |

Run Example 16: elementwise\_product\_example

|  |
| --- |
| python ./examples/src/main/python/ml/elementwise\_product\_example .py |

Operation results:

|  |
| --- |
| +-------------+-----------------+  | vector|transformedVector|  +-------------+-----------------+  |[1.0,2.0,3.0]| [0.0,2.0,6.0]|  |[4.0,5.0,6.0]| [0.0,5.0,12.0]|  +-------------+-----------------+ |

Run Use Case 17: estimator\_transformer\_param\_example

|  |
| --- |
| python ./examples/src/main/python/ml/estimator\_transformer\_param\_example .py |

Operation results:

|  |
| --- |
| features=[-1.0,1.5,1.3], label=1.0 -> prob=[0.0570730499357254,0.9429269500642746], prediction=1.0  features=[3.0,2.0,-0.1], label=0.0 -> prob=[0.9238521956443227,0.07614780435567725], prediction=0.0  features=[0.0,2.2,-1.5], label=1.0 -> prob=[0.10972780286187782,0.8902721971381222], prediction=1.0 |

Run Example 18: feature\_hasher\_example

|  |
| --- |
| python ./examples/src/main/python/ml/feature\_hasher\_example .py |

Operation results:

|  |
| --- |
| +----+-----+---------+------+--------------------------------------------------------+  |real|bool |stringNum|string|features |  +----+-----+---------+------+--------------------------------------------------------+  |2.2 |true |1 |foo |(262144,[174475,247670,257907,262126],[2.2,1.0,1.0,1.0])|  |3.3 |false|2 |bar |(262144,[70644,89673,173866,174475],[1.0,1.0,1.0,3.3]) |  |4.4 |false|3 |baz |(262144,[22406,70644,174475,187923],[1.0,1.0,4.4,1.0]) |  |5.5 |false|4 |foo |(262144,[70644,101499,174475,257907],[1.0,1.0,5.5,1.0]) |  +----+-----+---------+------+--------------------------------------------------------+ |

Run Example 19: fm\_classifier\_example

|  |
| --- |
| python ./examples/src/main/python/ml/fm\_classifier\_example .py |

Operation results:

|  |
| --- |
| +----------+------------+--------------------+  |prediction|indexedLabel| features|  +----------+------------+--------------------+  | 1.0| 1.0|(692,[98,99,100,1...|  | 1.0| 1.0|(692,[123,124,125...|  | 1.0| 1.0|(692,[123,124,125...|  | 1.0| 1.0|(692,[124,125,126...|  | 1.0| 1.0|(692,[124,125,126...|  +----------+------------+--------------------+  only showing top 5 rows  Test set accuracy = 0.970588 |

Run Example 20: fm\_regressor\_example

|  |
| --- |
| python ./examples/src/main/python/ml/fm\_regressor\_example .py |

Operation results:

|  |
| --- |
| +--------------------+-----+--------------------+  | prediction|label| features|  +--------------------+-----+--------------------+  |-0.08171237569487175| 0.0|(692,[95,96,97,12...|  |0.012256880643183952| 0.0|(692,[123,124,125...|  |-0.06536745201001372| 0.0|(692,[123,124,125...|  |-0.07546060710301175| 0.0|(692,[123,124,125...|  | 0.05862023527150016| 0.0|(692,[124,125,126...|  +--------------------+-----+--------------------+  only showing top 5 rows  Root Mean Squared Error (RMSE) on test data = 0.132506 |

Run Example 21: fpgrowth\_example

|  |
| --- |
| python ./examples/src/main/python/ml/fpgrowth\_example .py |

Operation results:

|  |
| --- |
| +---------+----+  |items|freq|  +---------+----+  | [1]| 3|  | [2]| 3|  | [2, 1]| 3|  | [5]| 2|  | [5, 2]| 2|  |[5, 2, 1]| 2|  | [5, 1]| 2|  +---------+----+  +----------+----------+------------------+----+------------------+  |antecedent|consequent| confidence|lift| support|  +----------+----------+------------------+----+------------------+  | [1]| [2]| 1.0| 1.0| 1.0|  | [1]| [5]|0.6666666666666666| 1.0|0.6666666666666666|  | [5, 1]| [2]| 1.0| 1.0|0.6666666666666666|  | [2]| [1]| 1.0| 1.0| 1.0|  | [2]| [5]|0.6666666666666666| 1.0|0.6666666666666666|  | [2, 1]| [5]|0.6666666666666666| 1.0|0.6666666666666666|  | [5]| [2]| 1.0| 1.0|0.6666666666666666|  | [5]| [1]| 1.0| 1.0|0.6666666666666666|  | [5, 2]| [1]| 1.0| 1.0|0.6666666666666666|  +----------+----------+------------------+----+------------------+  +---+------------+----------+  | id| items| prediction|  +---+------------+----------+  | 0| [1, 2, 5]| []|  | 1|[1, 2, 3, 5]| []|  | 2| [1, 2]| [5]|  +---+------------+----------+ |

Run Example 22: gaussian\_mixture\_example

|  |
| --- |
| python ./examples/src/main/python/ml/gaussian\_mixture\_example .py |

Operation results:

|  |
| --- |
| Gaussians shown as a DataFrame:  +-------------------------------------------------------------+----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------+  |mean |cov |  +-------------------------------------------------------------+----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------+  |[0.10000000000001552,0.10000000000001552,0.10000000000001552]|0.00666666666806454 0.006666666666806454 0.00666666666806454 \n0.006666666666806454 0.006666666666806454 0.006666666666806454 \n0.006666666666806454 0.006666666666806454 0.0066666666666806454 |  |[9.099999999999984,9.099999999999984,9.099999999999984] |0.006666666666808728 0.006666666666808728 0.00666666666808728 \n0.006666666666808728 0.006666666666808728 0.006666666666808728 \n0.006666666666808728 0.006666666666808728 0.0066666666666808728 |  +-------------------------------------------------------------+----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------+ |

Run Example 23: generalized\_linear\_regression\_example

|  |
| --- |
| python ./examples/src/main/python/ml/elementwise\_product\_example .py |

Operation results:

|  |
| --- |
| Coefficients: [0.010541828081257212,0.800325310056095,-0.7845165541420371,2.3679887171421914,0.5010002089857577 ,1.1222351159753026,-0.2926824398623296,-0.4983717432321304,-0.6035797180675657,0.672555006718746]  Intercept: 0.14592176145232041  Coefficient Standard Errors: [0.7950428434287478, 0.8049713176546897, 0.7975916824772489, 0.8312649247659919, 0.7945436200517938, 0.8118992572197593, 0.7919506385542777, 0.7973378214726764, 0.8300714999626418, 0.7771333489686802, 0.463930109648428]  T Values: [0.013259446542269238, 0.9942283563442595, -0.9836067393599172, 2.848657084633759, 0.6305509179635714, 1.382234441029355, -0.3695715687490668, -0.6250446546128238, -0.7271418403049983, 0.865430633766112, 0.31453393176593286]  P Values: [0.989426199114056, 0.32060241580811044, 0.3257943227369877, 0.004575078538306521, 0.5286281628105467, 0.16752945248679119, 0.7118614002322872, 0.5322327097421431, 0.467486325282384, 0.3872259825794293, 0.753249430501097]  Dispersion: 105.60988356821714  NullDeviance: 53229.3654338832  Residual Degree Of Freedom Null: 500  Deviance: 51748.8429484264  Residual Degree of Freedom: 490  AIC: 3769.1895871765314  Deviance Residuals:  +-------------------+  | devianceResiduals|  +-------------------+  |-10.974359174246889|  | 0.8872320138420559|  |-4.596541837478908|  |-20.411667435019638|  |-10.270419345342642|  |-6.0156058956799905|  |-10.663939415849267|  | 2.1153960525024718|  | 3.9807132379137675|  |-17.225218272069533|  |-4.611647633532147|  |6.4176669407698546|  | 11.407137945300537|  |-20.70176540467664|  |-2.6837485405109676|  |-16.755494794232536|  |8.154668342638725|  |-1.435505798735885|  |-0.6435058688185704|  |-1.13802589316832|  +-------------------+  only showing top 20 rows |

Run Example 24: gradient\_boosted\_tree\_classifier\_example

|  |
| --- |
| python ./examples/src/main/python/ml/elementwise\_product\_example .py |

Operation results:

|  |
| --- |
| +----------+------------+--------------------+  |prediction|indexedLabel| features|  +----------+------------+--------------------+  | 1.0| 1.0|(692,[95,96,97,12...|  | 1.0| 1.0|(692,[123,124,125...|  | 1.0| 1.0|(692,[124,125,126...|  | 1.0| 1.0|(692,[124,125,126...|  | 1.0| 1.0|(692,[125,126,127...|  +----------+------------+--------------------+  only showing top 5 rows  Test Error = 0  GBTClassificationModel: uid = GBTClassifier\_4c9bd5ed6b7c, numTrees=10, numClasses=2, numFeatures=692 |

Run Example 25: gradient\_boosted\_tree\_regressor\_example

|  |
| --- |
| python ./examples/src/main/python/ml/elementwise\_product\_example .py |

Operation results:

|  |
| --- |
| +----------+-----+--------------------+  |prediction|label| features|  +----------+-----+--------------------+  | 0.0| 0.0|(692,[122,123,148...|  | 0.0| 0.0|(692,[124,125,126...|  | 0.0| 0.0|(692,[124,125,126...|  | 0.0| 0.0|(692,[126,127,128...|  | 0.0| 0.0|(692,[126,127,128...|  +----------+-----+--------------------+  only showing top 5 rows  Root Mean Squared Error (RMSE) on test data = 0.19245  GBTRegressionModel: uid=GBTRegressor\_d208e681bc2a, numTrees=10, numFeatures=692 |

Run Example 26: imputer\_example

|  |
| --- |
| python ./examples/src/main/python/ml/imputer\_example .py |

Operation results:

|  |
| --- |
| +---+---+-----+-----+  | a| b|out\_a|out\_b|  +---+---+-----+-----+  |1.0|NaN| 1.0| 4.0|  |2.0|NaN| 2.0| 4.0|  |NaN|3.0| 3.0| 3.0|  |4.0|4.0| 4.0| 4.0|  |5.0|5.0| 5.0| 5.0|  +---+---+-----+-----+ |

Run Example 27: index\_to\_string\_example

|  |
| --- |
| python ./examples/src/main/python/ml/index\_to\_string\_example .py |

Operation results:

|  |
| --- |
| Transformed string column 'category' to indexed column 'categoryIndex'  +---+--------+-------------+  | id|category|categoryIndex|  +---+--------+-------------+  | 0| a| 0.0|  | 1| b| 2.0|  | 2| c| 1.0|  | 3| a| 0.0|  | 4| a| 0.0|  | 5| c| 1.0|  +---+--------+-------------+  StringIndexer will store labels in output column metadata  Transformed indexed column 'categoryIndex' back to original string column 'originalCategory' using labels in metadata  +---+-------------+----------------+  | id|categoryIndex|originalCategory|  +---+-------------+----------------+  | 0| 0.0| a|  | 1| 2.0| b|  | 2| 1.0| c|  | 3| 0.0| a|  | 4| 0.0| a|  | 5| 1.0| c|  +---+-------------+----------------+ |

Run use case 28: interaction\_example

|  |
| --- |
| python ./examples/src/main/python/ml/interaction\_example .py |

Operation results:

|  |
| --- |
| +---+--------------+--------------+------------------------------------------------------+  |id1|vec1 |vec2 |interactedCol |  +---+--------------+--------------+------------------------------------------------------+  |1 |[1.0,2.0,3.0] |[8.0,4.0,5.0] |[8.0,4.0,5.0,16.0,8.0,10.0,24.0,12.0,15.0] |  |2 |[4.0,3.0,8.0] |[7.0,9.0,8.0] |[56.0,72.0,64.0,42.0,54.0,48.0,112.0,144.0,128.0] |  |3 |[6.0,1.0,9.0] |[2.0,3.0,6.0] |[36.0,54.0,108.0,6.0,9.0,18.0,54.0,81.0,162.0] |  |4 |[10.0,8.0,6.0]|[9.0,4.0,5.0] |[360.0,160.0,200.0,288.0,128.0,160.0,216.0,96.0,120.0]|  |5 |[9.0,2.0,7.0] |[10.0,7.0,3.0]|[450.0,315.0,135.0,100.0,70.0,30.0,350.0,245.0,105.0] |  |6 |[1.0,1.0,4.0] |[2.0,8.0,4.0] |[12.0,48.0,24.0,12.0,48.0,24.0,48.0,192.0,96.0] |  +---+--------------+--------------+------------------------------------------------------+ |

Run Example 29: isotonic\_regression\_example

|  |
| --- |
| python ./examples/src/main/python/ml/ isotonic\_regression\_example .py |

Operation results:

|  |
| --- |
| Boundaries in increasing order: [0.01,0.17,0.18,0.27,0.28,0.29,0.3,0.31,0.34,0.35,0.36,0.41,0.42,0.71,0.72,0.74 ,0.75,0.76,0.77,0.78,0.79,0.8,0.81,0.82,0.83,0.84,0.85,0.86,0.87,0.88,0.89,1.0] |

Run example 30: kmeans\_example

|  |
| --- |
| python ./examples/src/main/python/ml/kmeans\_example .py |

Operation results:

|  |
| --- |
| Silhouette with squared euclidean distance = 0.9997530305375207  Cluster Centers:  [9.1 9.1 9.1]  [0.1 0.1 0.1] |

Run example 31: lda\_example

|  |
| --- |
| python ./examples/src/main/python/ml/lda\_example .py |

Operation results:

|  |
| --- |
| The lower bound on the log likelihood of the entire corpus: -786.7100101988664  The upper bound on perplexity: 3.025807731534101 |

Run use case 32: linear\_regression\_with\_elastic\_net

|  |
| --- |
| python ./examples/src/main/python/ml/linear\_regression\_with\_elastic\_net .py |

Operation results:

|  |
| --- |
| Coefficients: [0.0,0.3229251667740594,-0.3438548034562219,1.915601702345841,0.05288058680386255, 0.765962720459771,0.0,-0.15105392669186676,-0.21587930360904645,0.2202536918881343]  Intercept: 0.15989368442397356  numIterations: 6  objectiveHistory: [0.49999999999999994, 0.4967620357443381, 0.49363616643404634, 0.4936351537897608, 0.4936351214177871, 0.49363512062528014, 0.4936351206216114] |

Run case 33: linearsvc

|  |
| --- |
| python ./examples/src/main/python/ml/linearsvc .py |

Operation results:

|  |
| --- |
| Intercept: 0.5232286178786085 |

Run Example 34: logistic\_regression\_summary\_example

|  |
| --- |
| python ./examples/src/main/python/ml/logistic\_regression\_summary\_example .py |

Operation results:

|  |
| --- |
| objectiveHistory:  0.6833149135741672  0.6661906127558116  0.6207433672479604  0.6131541253123871  0.6059149689952391  0.5923656241678249  0.5898233082838021  0.5868012627420285  0.584443205871914  0.5830790068041739  0.5807015754032355  +---+--------------------+  |FPR| TPR|  +---+--------------------+  |0.0| 0.0|  |0.0|0.017543859649122806|  |0.0| 0.03508771929824561|  |0.0| 0.05263157894736842|  |0.0| 0.07017543859649122|  |0.0| 0.08771929824561403|  |0.0| 0.10526315789473684|  |0.0| 0.12280701754385964|  |0.0| 0.14035087719298245|  |0.0| 0.15789473684210525|  |0.0| 0.17543859649122806|  |0.0| 0.19298245614035087|  |0.0| 0.21052631578947367|  |0.0| 0.22807017543859648|  |0.0| 0.24561403508771928|  |0.0| 0.2631578947368421|  |0.0| 0.2807017543859649|  |0.0| 0.2982456140350877|  |0.0| 0.3157894736842105|  |0.0| 0.3333333333333333|  +---+--------------------+  only showing top 20 rows  areaUnderROC: 1.0 |

Run use case 35: logistic\_regression\_with\_elastic\_net

|  |
| --- |
| python ./examples/src/main/python/ml/logistic\_regression\_with\_elastic\_net .py |

Operation results:

|  |
| --- |
| Intercept: -0.5991460286401467  Multinomial coefficients: 2 X 692 CSRMatrix  (0,272) 0.0001  (0,300) 0.0001  (0,350) -0.0002  (0,351) -0.0001  (0,378) -0.0003  (0,379) -0.0002  (0,405) -0.0002  (0,406) -0.0004  (0,407) -0.0002  (0,433) -0.0003  (0,434) -0.0005  (0,435) -0.0001  (0,456) 0.0  (0,461) -0.0002  (0,462) -0.0004  (0,483) 0.0001  ..  ..  Multinomial intercepts: [0.2750587585718058,-0.2750587585718058] |

Run use case 36: max\_abs\_scaler\_example

|  |
| --- |
| python ./examples/src/main/python/ml/max\_abs\_scaler\_example .py |

Operation results:

|  |
| --- |
| +--------------+--------------------+  | features| scaledFeatures|  +--------------+--------------------+  |[1.0,0.1,-8.0]|[0.25,0.010000000...|  |[2.0,1.0,-4.0]| [0.5,0.1,-0.5]|  |[4.0,10.0,8.0]| [1.0,1.0,1.0]|  +--------------+--------------------+ |

Run Example 37: min\_hash\_lsh\_example

|  |
| --- |
| python ./examples/src/main/python/ml/min\_hash\_lsh\_example .py |

Operation results:

|  |
| --- |
| Features scaled to range: [0.000000, 1.000000]  +--------------+--------------+  | features|scaledFeatures|  +--------------+--------------+  |[1.0,0.1,-1.0]| (3,[],[])|  | [2.0,1.1,1.0]| [0.5,0.1,0.5]|  |[3.0,10.1,3.0]| [1.0,1.0,1.0]|  +--------------+--------------+ |

Run Example 38: min\_max\_scaler\_example

|  |
| --- |
| python ./examples/src/main/python/ml/min\_max\_scaler\_example .py |

Operation results:

|  |
| --- |
| Features scaled to range: [0.000000, 1.000000]  +--------------+--------------+  | features|scaledFeatures|  +--------------+--------------+  |[1.0,0.1,-1.0]| (3,[],[])|  | [2.0,1.1,1.0]| [0.5,0.1,0.5]|  |[3.0,10.1,3.0]| [1.0,1.0,1.0]|  +--------------+--------------+ |

Run Example 39: multiclass\_logistic\_regression\_with\_elastic\_net

|  |
| --- |
| python ./examples/src/main/python/ml/multiclass\_logistic\_regression\_with\_elastic\_net .py |

Operation results:

|  |
| --- |
| Coefficients:  3 X 4 CSR Matrix  (0,3) 0.305  (1,2) -0.7666  (1,3) -0.3854  Intercept: [0.0519258002072883,-0.12619173083598834,0.07426593062870003]  objectiveHistory:  1.098612288668108  1.0767872580801818  1.0324898663050683  1.0276148685544233  1.0292979666409194  1.0238563624020458  1.0236260835897757  1.0235478802964155  1.0231925082158746  1.0231565244620069  1.0229939986213705  False positive rate by label:  label 0: 0.22  label 1: 0.05  label 2: 0.0  True positive rate by label:  label 0: 1.0  label 1: 1.0  label 2: 0.46  Precision by label:  label 0: 0.6944444444444444  label 1: 0.9090909090909091  label 2: 1.0  Recall by label:  label 0: 1.0  label 1: 1.0  label 2: 0.46  F-measure by label:  label 0: 0.819672131147541  label 1: 0.9523809523809523  label 2: 0.6301369863013699  Accuracy: 0.82  FPR: 0.09  TPR: 0.82  F-measure: 0.800730023276621  Precision: 0.8678451178451179  Recall: 0.82 |

Run use case 40: multilayer\_perceptron\_classification

|  |
| --- |
| python ./examples/src/main/python/ml/multilayer\_perceptron\_classification .py |

Operation results:

|  |
| --- |
| Test set accuracy = 0.9523809523809523 |

Run Example 41: n\_gram\_example

|  |
| --- |
| python ./examples/src/main/python/ml/n\_gram\_example .py |

Operation results:

|  |
| --- |
| +----------------------------------------------------------------------------------+  |ngrams |  +----------------------------------------------------------------------------------+  |[Hi I, I heard, heard about, about Spark] |  |[I wish, wish Java, Java could, could use, use case, case classes]|  |[Logistic regression, regression models, models are, are neat] |  +----------------------------------------------------------------------------------+ |

Run Example 42: naive\_bayes\_example

|  |
| --- |
| python ./examples/src/main/python/ml/naive\_bayes\_example .py |

Operation results:

|  |
| --- |
| +-----+--------------------+--------------------+-----------+----------+  |label| features| rawPrediction|probability|prediction|  +-----+--------------------+--------------------+-----------+----------+  | 0.0|(692,[95,96,97,12...|[-172664.79564650...| [1.0,0.0]| 0.0|  | 0.0|(692,[98,99,100,1...|[-176279.15054306...| [1.0,0.0]| 0.0|  | 0.0|(692,[122,123,124...|[-189600.55409526...| [1.0,0.0]| 0.0|  | 0.0|(692,[124,125,126...|[-274673.88337431...| [1.0,0.0]| 0.0|  | 0.0|(692,[124,125,126...|[-183393.03869049...| [1.0,0.0]| 0.0|  | 0.0|(692,[125,126,127...|[-256992.48807619...| [1.0,0.0]| 0.0|  | 0.0|(692,[126,127,128...|[-210411.53649773...| [1.0,0.0]| 0.0|  | 0.0|(692,[127,128,129...|[-170627.63616681...| [1.0,0.0]| 0.0|  | 0.0|(692,[127,128,129...|[-212157.96750469...| [1.0,0.0]| 0.0|  | 0.0|(692,[127,128,129...|[-183253.80108550...| [1.0,0.0]| 0.0|  | 0.0|(692,[128,129,130...|[-246528.93739632...| [1.0,0.0]| 0.0|  | 0.0|(692,[150,151,152...|[-158348.34683571...| [1.0,0.0]| 0.0|  | 0.0|(692,[152,153,154...|[-210229.50765957...| [1.0,0.0]| 0.0|  | 0.0|(692,[152,153,154...|[-242985.16248889...| [1.0,0.0]| 0.0|  | 0.0|(692,[152,153,154...|[-94622.933454005...| [1.0,0.0]| 0.0|  | 0.0|(692,[153,154,155...|[-266465.39689814...| [1.0,0.0]| 0.0|  | 0.0|(692,[153,154,155...|[-144989.71469229...| [1.0,0.0]| 0.0|  | 0.0|(692,[154,155,156...|[-283834.57437738...| [1.0,0.0]| 0.0|  | 0.0|(692,[181,182,183...|[-155256.59399829...| [1.0,0.0]| 0.0|  | 1.0|(692,[100,101,102...|[-147726.11958982...| [0.0,1.0]| 1.0|  +-----+--------------------+--------------------+-----------+----------+  only showing top 20 rows  Test set accuracy = 1.0 |

Run example 43: normalizer\_example

|  |
| --- |
| python ./examples/src/main/python/ml/normalizer\_example .py |

Operation results:

|  |
| --- |
| Normalized using L^1 norm  +---+--------------+------------------+  | id| features| normFeatures|  +---+--------------+------------------+  | 0|[1.0,0.5,-1.0]| [0.4,0.2,-0.4]|  | 1| [2.0,1.0,1.0]| [0.5,0.25,0.25]|  | 2|[4.0,10.0,2.0]|[0.25,0.625,0.125]|  +---+--------------+------------------+  Normalized using L^inf norm  +---+--------------+--------------+  | id| features| normFeatures|  +---+--------------+--------------+  | 0|[1.0,0.5,-1.0]|[1.0,0.5,-1.0]|  | 1| [2.0,1.0,1.0]| [1.0,0.5,0.5]|  | 2|[4.0,10.0,2.0]| [0.4,1.0,0.2]|  +---+--------------+--------------+ |

Run use case 44: one\_vs\_rest\_example

|  |
| --- |
| python ./examples/src/main/python/ml/one\_vs\_rest\_example .py |

Operation results:

|  |
| --- |
| Test Error = 0.0444444 |

Run Example 45: onehot\_encoder\_example

|  |
| --- |
| python ./examples/src/main/python/ml/onehot\_encoder\_example .py |

Operation results:

|  |
| --- |
| +--------------+--------------+-------------+-------------+  |categoryIndex1|categoryIndex2| categoryVec1| categoryVec2|  +--------------+--------------+-------------+-------------+  | 0.0| 1.0|(2,[0],[1.0])|(2,[1],[1.0])|  | 1.0| 0.0|(2,[1],[1.0])|(2,[0],[1.0])|  | 2.0| 1.0| (2,[],[])|(2,[1],[1.0])|  | 0.0| 2.0|(2,[0],[1.0])| (2,[],[])|  | 0.0| 1.0|(2,[0],[1.0])|(2,[1],[1.0])|  | 2.0| 0.0| (2,[],[])|(2,[0],[1.0])|  +--------------+--------------+-------------+-------------+ |

Run Example 46: pca\_example

|  |
| --- |
| python ./examples/src/main/python/ml /pc a\_example .py |

Operation results:

|  |
| --- |
| +----------------------------------------------------------+  |pcaFeatures|  +----------------------------------------------------------+  |[1.6485728230883814,-4.013282700516299,0.4342930137624681]|  |[-4.645104331781533,-1.116797266361906,0.4342930137624673]|  |[-6.428880535676489,-5.33795142777536,0.4342930137624671] |  +----------------------------------------------------------+ |

Run use case 47: pipeline\_example

|  |
| --- |
| python ./examples/src/main/python/ml/pipeline\_example .py |

Operation results:

|  |
| --- |
| (4, spark ijk) --> prob=[0.6292098489668481,0.37079015103315194], prediction=0.000000  (5, lmn) --> prob=[0.984770006762304,0.015229993237696027], prediction=0.000000  (6, spark hadoop spark) --> prob=[0.13412348342566138,0.8658765165743386], prediction=1.000000  (7, apache hadoop) --> prob=[0.9955732114398529,0.00442678856014711], prediction=0.000000 |

Run Example 48: polynomial\_expansion\_example

|  |
| --- |
| python ./examples/src/main/python/ml/polynomial\_expansion\_example .py |

Operation results:

|  |
| --- |
| +----------+-----------------------------------------+  |features |polyFeatures |  +----------+-----------------------------------------+  |[2.0,1.0] |[2.0,4.0,8.0,1.0,2.0,4.0,1.0,2.0,1.0] |  |[0.0,0.0] |[0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0] |  |[3.0,-1.0]|[3.0,9.0,27.0,-1.0,-3.0,-9.0,1.0,3.0,-1.0]|  +----------+-----------------------------------------+ |

Run use case 49: power\_iteration\_clustering\_example

|  |
| --- |
| python ./examples/src/main/python/ml/power\_iteration\_clustering\_example .py |

Operation results:

|  |
| --- |
| +---+-------+  | id|cluster|  +---+-------+  | 0| 0|  | 1| 0|  | 2| 0|  | 3| 1|  | 4| 1|  +---+-------+ |

Run example 50: prefixspan\_example

|  |
| --- |
| python ./examples/src/main/python/ml/elementwise\_product\_example .py |

Operation results:

|  |
| --- |
| +----------+----+  |sequence|freq|  +----------+----+  | [[3]]| 2|  | [[1]]| 3|  | [[2]]| 3|  | [[1, 2]]| 3|  |[[1], [3]]| 2|  +----------+----+ |

Run Example 51: quantile\_discretizer\_example

|  |
| --- |
| python ./examples/src/main/python/ml/quantile\_discretizer\_example .py |

Operation results:

|  |
| --- |
| +---+----+------+  | id|hour|result|  +---+----+------+  | 0|18.0| 2.0|  | 1|19.0| 2.0|  | 2| 8.0| 1.0|  | 3| 5.0| 1.0|  | 4| 2.2| 0.0|  +---+----+------+ |

Run Example 52: random\_forest\_classifier\_example

|  |
| --- |
| python ./examples/src/main/python/ml/random\_forest\_classifier\_example .py |

Operation results:

|  |
| --- |
| +--------------+-----+--------------------+  |predictedLabel|label| features|  +--------------+-----+--------------------+  | 0.0| 0.0|(692,[95,96,97,12...|  | 0.0| 0.0|(692,[123,124,125...|  | 0.0| 0.0|(692,[124,125,126...|  | 0.0| 0.0|(692,[126,127,128...|  | 0.0| 0.0|(692,[126,127,128...|  +--------------+-----+--------------------+  only showing top 5 rows  Test Error = 0  RandomForestClassificationModel: uid=RandomForestClassifier\_2ccb1eaacf4f, numTrees=10, numClasses=2, numFeatures=692 |

Run Example 53: random\_forest\_regressor\_example

|  |
| --- |
| python ./examples/src/main/python/ml/random\_forest\_regressor\_example .py |

Operation results:

|  |
| --- |
| +----------+-----+--------------------+  |prediction|label| features|  +----------+-----+--------------------+  | 0.0| 0.0|(692,[98,99,100,1...|  | 0.05| 0.0|(692,[122,123,148...|  | 0.0| 0.0|(692,[123,124,125...|  | 0.0| 0.0|(692,[124,125,126...|  | 0.0| 0.0|(692,[124,125,126...|  +----------+-----+--------------------+  only showing top 5 rows  Root Mean Squared Error (RMSE) on test data = 0.0609071  RandomForestRegressionModel: uid=RandomForestRegressor\_317d813718fc, numTrees=20, numFeatures=692 |

Run Example 54: rformula\_example

|  |
| --- |
| python ./examples/src/main/python/ml/elementwise\_product\_example .py |

Operation results:

|  |
| --- |
| +--------------+-----+  | features|label|  +--------------+-----+  |[0.0,0.0,18.0]| 1.0|  |[1.0,0.0,12.0]| 0.0|  |[0.0,1.0,15.0]| 0.0|  +--------------+-----+ |

Run use case 55: robust\_scaler\_example

|  |
| --- |
| python ./examples/src/main/python/ml/robust\_scaler\_example .py |

Operation results:

|  |
| --- |
| +-----+--------------------+--------------------+  |label| features| scaledFeatures|  +-----+--------------------+--------------------+  | 0.0|(692,[127,128,129...|(692,[127,128,129...|  | 1.0|(692,[158,159,160...|(692,[158,159,160...|  | 1.0|(692,[124,125,126...|(692,[124,125,126...|  | 1.0|(692,[152,153,154...|(692,[152,153,154...|  | 1.0|(692,[151,152,153...|(692,[151,152,153...|  | 0.0|(692,[129,130,131...|(692,[129,130,131...|  | 1.0|(692,[158,159,160...|(692,[158,159,160...|  | 1.0|(692,[99,100,101,...|(692,[99,100,101,...|  | 0.0|(692,[154,155,156...|(692,[154,155,156...|  | 0.0|(692,[127,128,129...|(692,[127,128,129...|  | 1.0|(692,[154,155,156...|(692,[154,155,156...|  | 0.0|(692,[153,154,155...|(692,[153,154,155...|  | 0.0|(692,[151,152,153...|(692,[151,152,153...|  | 1.0|(692,[129,130,131...|(692,[129,130,131...|  | 0.0|(692,[154,155,156...|(692,[154,155,156...|  | 1.0|(692,[150,151,152...|(692,[150,151,152...|  | 0.0|(692,[124,125,126...|(692,[124,125,126...|  | 0.0|(692,[152,153,154...|(692,[152,153,154...|  | 1.0|(692,[97,98,99,12...|(692,[97,98,99,12...|  | 1.0|(692,[124,125,126...|(692,[124,125,126...|  +-----+--------------------+--------------------+  only showing top 20 rows |

Run case 56: sql\_transformer

|  |
| --- |
| python ./examples/src/main/python/ml/sql\_transformer .py |

Operation results:

|  |
| --- |
| +---+---+---+---+----+  | id| v1| v2| v3| v4|  +---+---+---+---+----+  | 0|1.0|3.0|4.0| 3.0|  | 2|2.0|5.0|7.0|10.0|  +---+---+---+---+----+ |

Run use case 57: standard\_scaler\_example

|  |
| --- |
| python ./examples/src/main/python/ml/standard\_scaler\_example .py |

Operation results:

|  |
| --- |
| +-----+--------------------+--------------------+  |label| features| scaledFeatures|  +-----+--------------------+--------------------+  | 0.0|(692,[127,128,129...|(692,[127,128,129...|  | 1.0|(692,[158,159,160...|(692,[158,159,160...|  | 1.0|(692,[124,125,126...|(692,[124,125,126...|  | 1.0|(692,[152,153,154...|(692,[152,153,154...|  | 1.0|(692,[151,152,153...|(692,[151,152,153...|  | 0.0|(692,[129,130,131...|(692,[129,130,131...|  | 1.0|(692,[158,159,160...|(692,[158,159,160...|  | 1.0|(692,[99,100,101,...|(692,[99,100,101,...|  | 0.0|(692,[154,155,156...|(692,[154,155,156...|  | 0.0|(692,[127,128,129...|(692,[127,128,129...|  | 1.0|(692,[154,155,156...|(692,[154,155,156...|  | 0.0|(692,[153,154,155...|(692,[153,154,155...|  | 0.0|(692,[151,152,153...|(692,[151,152,153...|  | 1.0|(692,[129,130,131...|(692,[129,130,131...|  | 0.0|(692,[154,155,156...|(692,[154,155,156...|  | 1.0|(692,[150,151,152...|(692,[150,151,152...|  | 0.0|(692,[124,125,126...|(692,[124,125,126...|  | 0.0|(692,[152,153,154...|(692,[152,153,154...|  | 1.0|(692,[97,98,99,12...|(692,[97,98,99,12...|  | 1.0|(692,[124,125,126...|(692,[124,125,126...|  +-----+--------------------+--------------------+  only showing top 20 rows |

Run example 58: stopwords\_remover\_example

|  |
| --- |
| python ./examples/src/main/python/ml/stopwords\_remover\_example .py |

Operation results:

|  |
| --- |
| +---+----------------------------+--------------------+  |id |raw |filtered |  +---+----------------------------+--------------------+  |0 |[I, saw, the, red, balloon] |[saw, red, balloon] |  |1 |[Mary, had, a, little, lamb]|[Mary, little, lamb]|  +---+----------------------------+--------------------+ |

Run Example 59: string\_indexer\_example

|  |
| --- |
| python ./examples/src/main/python/ml/string\_indexer\_example .py |

Operation results:

|  |
| --- |
| +---+--------+-------------+  | id|category|categoryIndex|  +---+--------+-------------+  | 0| a| 0.0|  | 1| b| 2.0|  | 2| c| 1.0|  | 3| a| 0.0|  | 4| a| 0.0|  | 5| c| 1.0|  +---+--------+-------------+ |

Run example 60: summarizer\_example

|  |
| --- |
| python ./examples/src/main/python/ml/summarizer\_example .py |

Operation results:

|  |
| --- |
| +-----+--------------------+  |label| features|  +-----+--------------------+  | 0.0|(20,[6,8,13,16],[...|  | 0.0|(20,[0,2,7,13,15,...|  | 1.0|(20,[3,4,6,11,19]...|  +-----+--------------------+ |

Run Example 61: tf\_idf\_example

|  |
| --- |
| python ./examples/src/main/python/ml/tf\_idf\_example .py |

Operation results:

|  |
| --- |
| +-----+--------------------+  |label| features|  +-----+--------------------+  | 0.0|(20,[6,8,13,16],[...|  | 0.0|(20,[0,2,7,13,15,...|  | 1.0|(20,[3,4,6,11,19]...|  +-----+--------------------+ |

Run Example 62: tokenizer\_example

|  |
| --- |
| python ./examples/src/main/python/ml/tokenizer\_example.py​​​ |

Operation results:

|  |
| --- |
| +-----------------------------------+------------------------------------------+-----+  |sentence |words |tokens|  +-----------------------------------+------------------------------------------+-----+  |Hi I heard about Spark |[hi, i, heard, about, spark] |5 |  |I wish Java could use case classes |[i, wish, java, could, use, case, classes]|7 |  |Logistic,regression,models,are,neat|[logistic,regression,models,are,neat] |1 |  +-----------------------------------+------------------------------------------+-----+  +-----------------------------------+------------------------------------------+-----+  |sentence |words |tokens|  +-----------------------------------+------------------------------------------+-----+  |Hi I heard about Spark |[hi, i, heard, about, spark] |5 |  |I wish Java could use case classes |[i, wish, java, could, use, case, classes]|7 |  |Logistic, regression, models, are, neat|[logistic, regression, models, are, neat] |5 |  +-----------------------------------+------------------------------------------+-----+ |

Run case 63: train\_validation\_split

|  |
| --- |
| python ./examples/src/main/python/ml/train\_validation\_split .py |

Operation results:

|  |
| --- |
| +--------------------+--------------------+--------------------+  | features| label| prediction|  +--------------------+--------------------+--------------------+  |(10,[0,1,2,3,4,5,...| -17.026492264209548| -1.780062242348691|  |(10,[0,1,2,3,4,5,...| -16.71909683360509| -0.1893325701092588|  |(10,[0,1,2,3,4,5,...| -15.375857723312297| 0.7252323736487188|  |(10,[0,1,2,3,4,5,...| -13.772441561702871| 3.2696413241677718|  |(10,[0,1,2,3,4,5,...| -13.039928064104615| 0.18817684046065775|  |(10,[0,1,2,3,4,5,...| -9.42898793151394| -3.449987079269568|  |(10,[0,1,2,3,4,5,...| -9.2679651250406|-0.33109075490696316|  |(10,[0,1,2,3,4,5,...| -9.173693798406978|-0.42727135281551937|  |(10,[0,1,2,3,4,5,...| -7.1500991588127265| 2.936884251408867|  |(10,[0,1,2,3,4,5,...| -6.930603551528371|-0.02839768193150...|  |(10,[0,1,2,3,4,5,...| -6.456944198081549| -0.9224776887934015|  |(10,[0,1,2,3,4,5,...| -3.2843694575334834| -1.0821208483033875|  |(10,[0,1,2,3,4,5,...| -1.99891354174786| 0.8052068273813595|  |(10,[0,1,2,3,4,5,...| -0.4683784136986876| 0.5046267770459569|  |(10,[0,1,2,3,4,5,...|-0.44652227528840105|0.053072145020589406|  |(10,[0,1,2,3,4,5,...| 0.10157453780074743| -1.0931313634366817|  |(10,[0,1,2,3,4,5,...| 0.2105613019270259| 1.068404443672188|  |(10,[0,1,2,3,4,5,...| 2.1214592666251364| 0.10226191630980659|  |(10,[0,1,2,3,4,5,...| 2.8497179990245116| 1.1908709522287462|  |(10,[0,1,2,3,4,5,...| 3.980473021620311| 2.3611397922073025|  +--------------------+--------------------+--------------------+  only showing top 20 rows |

Run Example 64: univariate\_feature\_selector\_example

|  |
| --- |
| python ./examples/src/main/python/ml/univariate\_feature\_selector\_example .py |

Operation results:

|  |
| --- |
| UnivariateFeatureSelector output with top 1 features selected using f\_classif  +---+--------------------+-----+----------------+  | id| features|label|selectedFeatures|  +---+--------------------+-----+----------------+  | 1|[1.7,4.4,7.6,5.8,...| 3.0| [2.3]|  | 2|[8.8,7.3,5.7,7.3,...| 2.0| [4.1]|  | 3|[1.2,9.5,2.5,3.1,...| 3.0| [2.5]|  | 4|[3.7,9.2,6.1,4.1,...| 2.0| [3.8]|  | 5|[8.9,5.2,7.8,8.3,...| 4.0| [3.0]|  | 6|[7.9,8.5,9.2,4.0,...| 4.0| [2.1]|  +---+--------------------+-----+----------------+ |

Run Example 65: variance\_threshold\_selector\_example

|  |
| --- |
| python ./examples/src/main/python/ml/variance\_threshold\_selector\_example .py |

Operation results:

|  |
| --- |
| Output: Features with variance lower than 8.000000 are removed.  +---+--------------------+-----------------+  | id| features| selectedFeatures|  +---+--------------------+-----------------+  | 1|[6.0,7.0,0.0,7.0,...|[6.0,0.0,7.0,0.0]|  | 2|[0.0,9.0,6.0,0.0,...|[0.0,6.0,0.0,9.0]|  | 3|[0.0,9.0,3.0,0.0,...|[0.0,3.0,0.0,5.0]|  | 4|[0.0,9.0,8.0,5.0,...|[0.0,8.0,5.0,4.0]|  | 5|[8.0,9.0,6.0,5.0,...|[8.0,6.0,5.0,4.0]|  | 6|[8.0,9.0,6.0,0.0,...|[8.0,6.0,0.0,0.0]|  +---+--------------------+-----------------+ |

Run example 66: vector\_assembler\_example

|  |
| --- |
| python ./examples/src/main/python/ml/vector\_assembler\_example .py |

Operation results:

|  |
| --- |
| Assembled columns 'hour', 'mobile', 'userFeatures' to vector column 'features'  +-----------------------+-------+  |features |clicked|  +-----------------------+-------+  |[18.0,1.0,0.0,10.0,0.5]|1.0 |  +-----------------------+-------+ |

Run Example 67: vector\_indexer\_example

|  |
| --- |
| python ./examples/src/main/python/ml/vector\_indexer\_example .py |

Operation results:

|  |
| --- |
| +-----+--------------------+--------------------+  |label| features| indexed|  +-----+--------------------+--------------------+  | 0.0|(692,[127,128,129...|(692,[127,128,129...|  | 1.0|(692,[158,159,160...|(692,[158,159,160...|  | 1.0|(692,[124,125,126...|(692,[124,125,126...|  | 1.0|(692,[152,153,154...|(692,[152,153,154...|  | 1.0|(692,[151,152,153...|(692,[151,152,153...|  | 0.0|(692,[129,130,131...|(692,[129,130,131...|  | 1.0|(692,[158,159,160...|(692,[158,159,160...|  | 1.0|(692,[99,100,101,...|(692,[99,100,101,...|  | 0.0|(692,[154,155,156...|(692,[154,155,156...|  | 0.0|(692,[127,128,129...|(692,[127,128,129...|  | 1.0|(692,[154,155,156...|(692,[154,155,156...|  | 0.0|(692,[153,154,155...|(692,[153,154,155...|  | 0.0|(692,[151,152,153...|(692,[151,152,153...|  | 1.0|(692,[129,130,131...|(692,[129,130,131...|  | 0.0|(692,[154,155,156...|(692,[154,155,156...|  | 1.0|(692,[150,151,152...|(692,[150,151,152...|  | 0.0|(692,[124,125,126...|(692,[124,125,126...|  | 0.0|(692,[152,153,154...|(692,[152,153,154...|  | 1.0|(692,[97,98,99,12...|(692,[97,98,99,12...|  | 1.0|(692,[124,125,126...|(692,[124,125,126...|  +-----+--------------------+--------------------+  only showing top 20 rows |

Run Example 68: vector\_size\_hint\_example

|  |
| --- |
| python ./examples/src/main/python/ml/vector\_size\_hint\_example .py |

Operation results:

|  |
| --- |
| Rows where 'userFeatures' is not the right size are filtered out  +---+----+------+--------------+-------+  |id |hour|mobile|userFeatures |clicked|  +---+----+------+--------------+-------+  |0 |18 |1.0 |[0.0,10.0,0.5]|1.0 |  +---+----+------+--------------+-------+  Assembled columns 'hour', 'mobile', 'userFeatures' to vector column 'features'  +-----------------------+-------+  |features |clicked|  +-----------------------+-------+  |[18.0,1.0,0.0,10.0,0.5]|1.0 |  +-----------------------+-------+ |

Run Example 69: vector\_slicer\_example

|  |
| --- |
| python ./examples/src/main/python/ml/vector\_slicer\_example .py |

Operation results:

|  |
| --- |
| +--------------------+-------------+  |userFeatures| features|  +--------------------+-------------+  |(3,[0,1],[-2.0,2.3])|(1,[0],[2.3])|  | [-2.0,2.3,0.0]| [2.3]|  +--------------------+-------------+ |

Run Example 70: word2vec\_example

|  |
| --- |
| python ./examples/src/main/python/ml/word2vec\_example .py |

Operation results:

|  |
| --- |
| Text: [Hi, I, heard, about, Spark] =>  Vector: [0.032865070179104804,0.002630169689655304,0.02410868888255209]  Text: [I, wish, Java, could, use, case, classes] =>  Vector: [-0.025600749999284744,-0.08642657420464923,0.02152669057250023]  Text: [Logistic, regression, models, are, neat] =>  Vector: [0.04076812872663141,0.045929113030433656,0.05449362322688103] |

1. Python test case 2: test case in mllib/folder

Run Example 71 : binary\_classification\_metrics\_example.py

|  |
| --- |
| # Note that it is best to enter the spark-3.5.1 folder to run the test case  python ./examples/src/main/python/ml lib /binary\_classification\_metrics\_example .py |

Operation results:

|  |
| --- |
| Area under PR = 1.0  Area under ROC = 1.0 |

Run example 72: mllib\_bisecting\_k\_means\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /bisecting\_k\_means\_example .py |

Operation results:

|  |
| --- |
| Bisecting K-means Cost = 0.11999999999999958 |

Run case 73: correlations

|  |
| --- |
| python ./examples/src/main/python/ml lib /correlations .py |

Operation results:

|  |
| --- |
| Summary of data file: data/mllib/sample\_linear\_regression\_data.txt  501 data points  Correlation (pearson) between label and each feature  Feature Correlation  0 0.00595645  1 0.0332056  2 -0.0406646  3 0.123178  4 0.0240118  5 0.0648617  6 -0.0223995  7 -0.0279813  8 -0.0359889  9 0.0345207 |

Run example 74: correlations\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /correlations\_example .py |

Operation results:

|  |
| --- |
| Correlation is: 0.8500286768773005  [[1. 0.97888347 0.99038957]  [ 0.97888347 1 0.99774832 ]  [ 0.99038957 0.99774832 1. ]] |

Run example 75: mllib\_decision\_tree\_classification\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /decision\_tree\_classification\_example .py |

Operation results:

|  |
| --- |
| Test Error = 0.02857142857142857  Learned classification tree model:  DecisionTreeModel classifier of depth 2 with 5 nodes  If (feature 434 <= 70.5)  If (feature 99 <= 35.0)  Predict: 0.0  Else (feature 99 > 35.0)  Predict: 1.0  Else (feature 434 > 70.5)  Predict: 1.0 |

Run example 76: mllib\_decision\_tree\_regression\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /decision\_tree\_regression\_example .py |

Operation results:

|  |
| --- |
| Test Mean Squared Error = 0.04  Learned regression tree model:  DecisionTreeModel regressor of depth 2 with 5 nodes  If (feature 406 <= 126.5)  If (feature 100 <= 193.5)  Predict: 0.0  Else (feature 100 > 193.5)  Predict: 1.0  Else (feature 406 > 126.5)  Predict: 1.0 |

Run example 77: mllib\_elementwise\_product\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /elementwise\_product\_example .py |

Operation results:

|  |
| --- |
| transformedData:  [0.0,0.0,0.0]  [0.0,0.1,0.2]  [0.0,0.2,0.4]  [0.0,9.0,18.0]  [0.0,9.1,18.2]  [0.0,9.2,18.4]  transformedData2:  0.0  0.0  0.0 |

Run example 78: mllib\_fpgrowth\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /fpgrowth\_example .py |

Operation results:

|  |
| --- |
| FreqItemset(items=['z'], freq=5)  FreqItemset(items=['x'], freq=4)  FreqItemset(items=['x', 'z'], freq=3)  FreqItemset(items=['y'], freq=3)  FreqItemset(items=['y', 'x'], freq=3)  FreqItemset(items=['y', 'x', 'z'], freq=3) |

Run example 79: mllib\_gaussian\_mixture\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /gaussian\_mixture\_example .py |

Operation results:

|  |
| --- |
| # RISC-V output  weight = 0.4804289027506675 mu = [0.07222638933002806,0.016680304233958256] sigma = [[4.78069599 1.87671289]  [1.87671289 0.91484668]]  weight = 0.5195710972493325 mu = [-0.10443327095608487,0.042857091804250146] sigma = [[ 4.90617163 -2.00629664]  [-2.00629664 1.01125534]]  # ARM output  File "/home/zhuchenhao/spark-3.5.1/python/lib/py4j-0.10.9.7-src.zip/py4j/protocol.py", line 326, in get\_return\_value  py4j.protocol.Py4JJavaError: An error occurred while calling o33.save.  : org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory file:/home/zhuchenhao/spark-3.5.1/target/org/apache/spark/PythonGaussianMixtureExample/GaussianMixtureModel/metadata already exists |

According to the ARM error, there is already an output model in the output path, which needs to be deleted before running the use case:   
Delete file:

|  |
| --- |
| rm -rf spark-3.5.1/target/org/apache/spark/PythonGaussianMixtureExample/GaussianMixtureModel |

Operation results:

|  |
| --- |
| weight = 0.4804289027506675 mu = [0.07222638933002806,0.016680304233958256] sigma = [[4.78069599 1.87671289]  [1.87671289 0.91484668]]  weight = 0.5195710972493325 mu = [-0.10443327095608487,0.042857091804250146] sigma = [[ 4.90617163 -2.00629664]  [-2.00629664 1.01125534]] |

Run case 80: gaussian\_mixture\_model

|  |
| --- |
| python ./examples/src/main/python/ml lib /gaussian\_mixture\_model .py |

Operation results:

|  |
| --- |
| usage: gaussian\_mixture\_model.py [-h] [--convergenceTol CONVERGENCETOL]  [--maxIterations MAXITERATIONS] [--seed SEED]  inputFile k  gaussian\_mixture\_model.py: error: the following arguments are required: inputFile, k |

The number of data files and mixture components that need to be input:

Run the test case again

|  |
| --- |
| python ./examples/src/main/python/ml lib /gaussian\_mixture\_model .py data/mllib/gmm\_data.txt 2 |

Operation results:

|  |
| --- |
| ValueError: could not convert string to float: '' |

Run Example 81: gradient\_boosting\_classification\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /gradient\_boosting\_classification\_example .py |

Operation results:

|  |
| --- |
| Test Error = 0.03571428571428571  Learned classification GBT model:  TreeEnsembleModel classifier with 3 trees  Tree 0:  If (feature 406 <= 147.5)  If (feature 100 <= 193.5)  Predict: -1.0  Else (feature 100 > 193.5)  Predict: 1.0  Else (feature 406 > 147.5)  Predict: 1.0  Tree 1:  If (feature 462 <= 62.5)  If (feature 379 <= 251.5)  Predict: -0.47681168808847024  Else (feature 379 > 251.5)  Predict: 0.4768116880884694  Else (feature 462 > 62.5)  Predict: 0.4768116880884701  Tree 2:  If (feature 434 <= 124.5)  If (feature 379 <= 251.5)  Predict: -0.4381935810427206  Else (feature 379 > 251.5)  Predict: 0.43819358104271977  Else (feature 434 > 124.5)  Predict: 0.43819358104272055 |

Run example 82: gradient\_boosting\_regression\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /gradient\_boosting\_regression\_example .py |

Operation results:

|  |
| --- |
| Test Mean Squared Error = 0.02702702702702703  Learned regression GBT model:  TreeEnsembleModel regressor with 3 trees  Tree 0:  If (feature 434 <= 70.5)  Predict: 0.0  Else (feature 434 > 70.5)  Predict: 1.0  Tree 1:  Predict: 0.0  Tree 2:  Predict: 0.0 |

Run example 83: hypothesis\_testing\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /hypothesis\_testing\_example .py |

Operation results:

|  |
| --- |
| Chi squared test summary:  method: pearson  degrees of freedom = 4  statistic = 0.12499999999999999  pValue = 0.998126379239318  No presumption against null hypothesis: observed follows the same distribution as expected..  Chi squared test summary:  method: pearson  degrees of freedom = 2  statistic = 0.14141414141414144  pValue = 0.931734784568187  No presumption against null hypothesis: the occurrence of the outcomes is statistically independent..  Column 1:  Chi squared test summary:  method: pearson  degrees of freedom = 0  statistic = 0.0  pValue = 1.0  No presumption against null hypothesis: the occurrence of the outcomes is statistically independent..  Column 2:  Chi squared test summary:  method: pearson  degrees of freedom = 0  statistic = 0.0  pValue = 1.0  No presumption against null hypothesis: the occurrence of the outcomes is statistically independent..  Column 3:  Chi squared test summary:  method: pearson  degrees of freedom = 0  statistic = 0.0  pValue = 1.0  No presumption against null hypothesis: the occurrence of the outcomes is statistically independent.. |

Run example 84: hypothesis\_testing\_kolmogorov\_smirnov\_test\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /hypothesis\_testing\_kolmogorov\_smirnov\_test\_example .py |

Operation results:

|  |
| --- |
| Kolmogorov-Smirnov test summary:  degrees of freedom = 0  statistic = 0.539827837277029  pValue = 0.06821463111921133  Low presumption against null hypothesis: Sample follows theoretical distribution. |

Run example 85: isotonic\_regression\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /isotonic\_regression\_example .py |

Operation results:

|  |
| --- |
| Mean Squared Error = 0.006929568245892503 |

Run Example 86: k\_means\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /k\_means\_example .py |

Operation results:

|  |
| --- |
| Within Set Sum of Squared Error = 0.6928203230275529 |

Run example 87: kernel\_density\_estimation\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /kernel\_density\_estimation\_example .py |

Operation results:

|  |
| --- |
| [ 0.04145944 0.07902017 0.0896292 ] |

Run case 88: kmean.py

|  |
| --- |
| python ./examples/src/main/python/ml lib / kmean.py |

Operation results:

|  |
| --- |
| Usage: kmeans <file> <k> |

This use case requires input data and specifies the number of clusters.

Specify k=3 and re-run the test case:

|  |
| --- |
| python examples/src/main/python/mllib/kmeans.py data/mllib/kmeans\_data.txt 3 |

Operation results:

|  |
| --- |
| Final centers: [array([9.1, 9.1, 9.1]), array([0.05, 0.05, 0.05]), array([0.2, 0.2, 0.2])]  Total Cost: 0.07499999999999957 |

Run Example 89: latent\_dirichlet\_allocation\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /latent\_dirichlet\_allocation\_example .py |

Operation results:

|  |
| --- |
| Learned topics (as distributions over vocab of 11 words):  Topic 0:  10.844874523176097  13.387466374041061  3.1641006356606125  1.604087654220152  16.854343606196224  16.811685579899144  5.890960214002804  8.486152140824787  1.5010672474092281  6.3899513088994215  2.106558651001079 |

Run example 90: linear\_regression\_with\_sgd\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /linear\_regression\_with\_sgd\_example .py |

Operation results:

|  |
| --- |
| Mean Squared Error = 7.4510328101026015 |

Run case 91: logistic\_regression

|  |
| --- |
| python ./examples/src/main/python/ml lib /logistic\_regression .py |

Operation results:

|  |
| --- |
| Usage: logistic\_regression <file> <iterations> |

This use case requires specifying the input data file and determining the number of iterations.

Rerun this test case:

Run the example: logistic\_regression

|  |
| --- |
| python ./examples/src/main/python/ml lib /logistic\_regression .py data/mllib/sample\_svm\_data.txt 10 |

Operation results:

|  |
| --- |
| Final weights: [-0.16663178606008336,0.004722240688635306,-0.05740572013019209,0.1765867182694578 ,-0.10851094748926805,0.0006737410115516748,0.13325284728567566,-0.2276230426720264 3,0.015687806873713177,0.015687806873713156,0.1660962863025941,-0.1259145555727007,0.12305566740265307,-0.07249567869896505,-0.031086230558764274,0.08306150811329616]  Final intercept: 0.0 |

Run example 92: logistic\_regression\_with\_lbfgs\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /logistic\_regression\_with\_lbfgs\_example .py |

Operation results:

|  |
| --- |
| Training Error = 0.36645962732919257 |

Run example 93: multi\_class\_metrics\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib / multi\_ class\_metrics\_example .py |

Operation results:

|  |
| --- |
| Summary Stats  Precision = 1.0  Recall = 1.0  F1 Score = 1.0  Class 0.0 precision = 0.6818181818181818  Class 0.0 recall = 0.9375  Class 0.0 F1 Measure = 0.7894736842105263  Class 1.0 precision = 1.0  Class 1.0 recall = 1.0  Class 1.0 F1 Measure = 1.0  Class 2.0 precision = 0.9285714285714286  Class 2.0 recall = 0.65  Class 2.0 F1 Measure = 0.7647058823529412  Weighted recall = 0.8461538461538461  Weighted precision = 0.8746253746253747  Weighted F(1) Score = 0.8447249345082164  Weighted F(0.5) Score = 0.8585331672376207  Weighted false positive rate = 0.0718482905982906 |

Run example 94: multi\_label\_metrics\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /multi\_label\_metrics\_example .py |

Operation results:

|  |
| --- |
| Recall = 0.6428571428571429  Precision = 0.6666666666666666  F1 measure = 0.6380952380952382  Accuracy = 0.5476190476190476  Class 0.0 precision = 1.0  Class 0.0 recall = 0.8  Class 0.0 F1 Measure = 0.888888888888889  Class 1.0 precision = 0.6666666666666666  Class 1.0 recall = 0.6666666666666666  Class 1.0 F1 Measure = 0.6666666666666666  Class 2.0 precision = 0.5  Class 2.0 recall = 0.5  Class 2.0 F1 Measure = 0.5  Micro precision = 0.7272727272727273  Micro recall = 0.6666666666666666  Micro F1 measure = 0.6956521739130435  Hamming loss = 0.33333333333333333  Subset accuracy = 0.2857142857142857 |

Run example 95: naive\_bayes\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /naive\_bayes\_example .py |

Operation results:

|  |
| --- |
| Model accuracy 1.0  sameModel accuracy 1.0 |

Run example 96: normalizer\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /normalizer\_example .py |

Operation results:

|  |
| --- |
| data1:  (0.0, SparseVector(692, {127: 0.0196, 128: 0.0612, 129: 0.0974, 130: 0.0612, 131: 0.0192, 154: 0.0185, 155: 0.0916, 156: 0.097, 157: 0.097, 158: 0.097, 159: 0.0912, 181: 0.0208, 182: 0.0874, 183: 0.0974, 184: 0.097, 185: 0.092, 186: 0.0897, 187: 0.097, 188: 0.0219, 218: 0.0023, 219: 0.0037, 220: 0.0046, 221: 0.0047, 222: 0.0046, 223: 0.0047, 224: 0.0046, 225: 0.097, 226: 0.097, 227: 0.097, 228: 0.097, 229: 0.0862, 230: 0.097, 231: 0.097, 0.0727, 272: 0.0974, 273: 0.0647, 289: 0.0185, 290: 0.0916, 291: 0.0974, 292: 0.0196, 293: 0.0916, 294: 0.0974, 295: 0.0974, 296: 0.0731, 297: 0.0439, 298: 0.0974, 299: 0.0878, 270: 0.0181, 271: 0.0304, 272: 0.0974, 273: 0.0647, 289: 0.0185, 290: 0.0916, 291: 0.097 318: 0.0974, 319: 0.0897, 320: 0.0801, 321: 0.0323, 328: 0.0974, 329: 0.097, 330: 0.0635, 343: 0.0027, 372: 0.097, 373: 0.097, 374: 0.0242, 384: 0.0974, 385: 0.097, 386: 0.0751, 399: 0.0762, 400: 0.0974, 401: 0.0731, 414: 0.0754, 426: 0.0293, 427: 0.0947, 428: 0.097, 429: 0.0431, 440: 0.0974, 441: 0.097, 442: 0.057, 454: 0.0327, 455: 0.097, 456: 0.0885, 457: 0.0096, 466: 0.0027, 467: 0.052, 468: 0.0974, 469: 0.0716, 470: 0.0046, 0.0543, 520: 0.0185, 521: 0.0635, 522: 0.097, 523: 0.0666, 538: 0.0331, 539: 0.0974, 540: 0.0866, 547: 0.0273, 0.0542, 570: 0.0185, 571: 0.0112, 572: 0.0327, 573: 0.0685, 574: 0.0866, 575: 0.0974, 576: 0.0858, 577: 0.0643, 578: 0.0216, 594: 0.0327, 595: 0.097, 596: 0.097, 0.097, 627: 0.097, 628: 0.097, 629: 0.0897, 630: 0.0558, 651: 0.0096, 652: 0.0493, 653: 0.0766, 654: 0.0571 0.097, 654: 0.0974, 655: 0.097, 656: 0.0543, 657: 0.0142})) |

Run Example 97: pca\_rowmatrix\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /pca\_rowmatrix\_example .py |

Operation results:

|  |
| --- |
| Projected Row Matrix of principal component:  [1.6485728230883814,-4.0132827005162985,-1.0091435193998504,-5.250587616986039]  [-4.645104331781532,-1.1167972663619048,-1.0091435193998501,-5.250587616986039]  [-6.428880535676488,-5.337951427775359,-1.009143519399851,-5.250587616986039] |

Run example 98: mllib\_power\_iteration\_clustering\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /power\_iteration\_clustering\_example .py |

Operation results:

|  |
| --- |
| 13 -> 1  15 -> 1  11 -> 1  1 -> 0  3 -> 0  7 -> 0  9 -> 0  5 -> 1  4 -> 0  14 -> 0  0 -> 0  6 -> 0  8 -> 0  12 -> 0  10 -> 0  2 -> 0 |

Run Example 99: random\_forest\_classification\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /random\_forest\_classification\_example .py |

Operation results:

|  |
| --- |
| Test Error = 0.030303030303030304  Learned classification forest model:  TreeEnsembleModel classifier with 3 trees  Tree 0:  If (feature 462 <= 62.5)  Predict: 0.0  Else (feature 462 > 62.5)  Predict: 1.0  Tree 1:  If (feature 455 <= 25.5)  Predict: 1.0  Else (feature 455 > 25.5)  Predict: 0.0  Tree 2:  If (feature 300 <= 32.0)  If (feature 434 <= 70.5)  If (feature 240 <= 143.0)  Predict: 0.0  Else (feature 240 > 143.0)  Predict: 1.0  Else (feature 434 > 70.5)  Predict: 1.0  Else (feature 300 > 32.0)  Predict: 0.0 |

Run example 100: random\_forest\_regression\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /random\_forest\_regression\_example .py |

Operation results:

|  |
| --- |
| Test Mean Squared Error = 0.05952380952380953  Learned regression forest model:  TreeEnsembleModel regressor with 3 trees  Tree 0:  If (feature 405 <= 31.5)  Predict: 0.0  Else (feature 405 > 31.5)  Predict: 1.0  Tree 1:  If (feature 490 <= 29.0)  Predict: 0.0  Else (feature 490 > 29.0)  Predict: 1.0  Tree 2:  If (feature 469 <= 4.0)  Predict: 1.0  Else (feature 469 > 4.0)  Predict: 0.0 |

Running Use Case 101: random\_rdd\_generation

|  |
| --- |
| python ./examples/src/main/python/ml lib /random\_rdd\_generation .py |

Operation results:

|  |
| --- |
| Generated RDD of 10000 examples sampled from the standard normal distribution  First 5 samples:  -0.2765753144483936  -0.3530067879647157  -1.0938016504592012  -0.6155908896400413  0.505877490414044  Generated RDD of 10000 examples of length-2 vectors.  First 5 samples:  [-0.02860506 1.72750605]  [-1.48931773 -0.56705206]  [-0.25451574 -2.30667555]  [ 0.60067764 1.38769365 ]  [ 0.90695625 1.6567046 ] |

Run use case 102: ranking\_metrics\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /ranking\_metrics\_example .py |

Operation results:

|  |
| --- |
| RMSE = 0.24439698505307364  R-squared = 0.9575988513437905 |

Run use case 103: recommendation\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /recommendation\_example .py |

Operation results:

|  |
| --- |
| Mean Squared Error = 5.5386021167071584e-06 |

Run use case 104: regression\_metrics\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /regression\_metrics\_example .py |

Operation results:

|  |
| --- |
| MSE = 103.30968681818085  RMSE = 10.164137288436281  R-squared = 0.027639110967836777  MAE = 8.148691907953307  Explained variance = 2.8883952017178944 |

Run use case 105: sampled\_rdds

|  |
| --- |
| python ./examples/src/main/python/ml lib /sampled\_rdds .py |

Operation results:

|  |
| --- |
| Loaded data with 100 examples from file: data/mllib/sample\_binary\_classification\_data.txt  Sampling RDD using fraction 0.1. Expected sample size = 10.  RDD.sample(): sample has 9 examples  RDD.takeSample(): sample has 10 examples  Keyed data using label (Int) as key ==> Orig  Sampled 16 examples using approximate stratified sampling (by label). ==> Sample  Fractions of examples with key  Key Orig Sample  0 0.43 0.4375  1 0.57 0.5625 |

Run use case 106: standard\_scaler\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib / s tandard\_scaler\_example .py |

Operation results:

|  |
| --- |
| data1:  (0.0, SparseVector(692, {127: 0.5468, 128: 1.5923, 129: 2.4354, 130: 1.7081, 131: 0.7335, 154: 0.4346, 155: 2.0985, 156: 2.2563, 157: 2.2368, 158: 2.2269, 159: 2.2555, 181: 0.4713, 182: 2.0575, 183: 2.3318, 184: 2.3761, 185: 2.1237, 186: 2.0452,187: 214: 0.7116, 215: 2.2256, 216: 2.4032, 217: 1.5931, 235: 1.5394, 236: 2.188, 237: 2.1493, 238: 2.2924, 239: 2.3889, 240: 267: 1.083, 268: 2.1786, 269: 2.0345, 270: 0.4392, 271: 0.7218, 272: 2.2177, 273: 1.6764, 289: 0.4794, 290: 318: 2.3539, 319: 2.1535, 320: 1.9834, 321: 0.8017, 328: 2.4941, 329: 2.3661, 330: 2.277, 331: 2.3083, 332: 0.5397, 333: 1.6059, 334: 2.3539, 335: 2.214, 336: 2.21 373: 2.3315, 374: 0.6414, 384: 2.5939, 385: 2.4196, 386: 1.8986, 399: 0.2749, 400: 0.591, 401: 2.4489, 402: 2.5746, 403: 2.4123, 404: 0.6411 44: 1.5716, 454: 0.8453, 455: 2.3533, 456: 2.1201, 457: 0.2628, 466: 0.0965, 467: 1.4473, 468: 2.5738, 469: 1.1435, 440: 2.5745, 441: 2.4723, 442: 1.5716, 520: 0.6477, 521: 1.9663, 522: 2.3879, 523: 1.6775, 538: 0.8901, 539: 2.3049, 540: 1.4911, 495: 2.4922, 496: 2.224, 497: 0.7444, 510: 0.8429, 511: 2.314, 512: 1.3481, 520: 0.6477, 521: 1.9663, 522: 2.3879, 523: 1.6775, 538: 0.8901, 539: 2.3049, 540: 574: 2.0206, 575: 2.349, 576: 2.0094, 577: 1.4659, 578: 2.5412, 594: 1.4063, 595: 624: 2.4617, 625: 2.2485, 626: 2.208, 627: 2.323, 628: 2.306, 629: 2.1166, 630: 1.2687, 651: 0.6799, 652: 2.7625, 653: 2.811, 654: 2.706 0.5922, 652: 1.7772, 653: 2.3485, 654: 2.1973, 655: 2.25, 656: 1.2522, 657: 0.3419})) |

Run Example 107: stratified\_sampling\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /stratified\_sampling\_example .py |

Operation results:

|  |
| --- |
| (2, 'd')  (3, 'f') |

Run use case 108: streaming\_k\_means\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /streaming\_k\_means\_example .py |

Operation results:

|  |
| --- |
| -------------------------------------------  Time: 2024-07-22 17:53:57  -------------------------------------------  (1.0, 0)  (2.0, 1)  -------------------------------------------  Time: 2024-07-22 17:53:58  -------------------------------------------  Final centers: [[ 4.19486462 4.00002246 4.08267685]  [ 2.2408932 1.86755799 -0.97727788 ]] |

Run use case 109: streaming\_linear\_regression\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /streaming\_linear\_regression\_example .py |

Operation results:

|  |
| --- |
| Usage: streaming\_linear\_regression\_example.py <trainingDir> <testDir> |

This use case requires specifying a training set and a test set. (The format of the training set and the test set is not known yet)

Run example 110: summary\_statistics\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /summary\_statistics\_example .py |

Operation results:

|  |
| --- |
| [2. 20. 200.]  [1.e+00 1.e+02 1.e+04]  [3. 3. 3.] |

Run example 111: svd\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /svd\_example .py |

Operation results:

|  |
| --- |
| U factor is:  [-0.38829130511665644,-0.9198099362554474,-0.056387441301709175,9.313225746154785e-09,0.0]  [-0.5301719995198351,0.2730185511901228,-0.8027319114319463,0.0,0.0]  [-0.7537556058139434,0.2817987790459642,0.5936682026454339,1.4901161193847656e-08,1.4901161193847656e-08]  Singular values are: [13.029275535600473,5.368578733451684,2.5330498218813755,6.323166049206486e-08,2.0226934557075942e-08]  V factor is:  DenseMatrix([[-0.31278534, 0.31167136, 0.30366911, 0.8409913, -0.07446478],  [-0.02980145, -0.17133211, -0.02226069, 0.14664984, 0.97352733],  [-0.12207248, 0.15256471, -0.95070998, 0.23828799, -0.03452092],  [-0.71847899, -0.68096285, -0.0172245, -0.02094998, -0.13907533],  [-0.60841059, 0.62170723, 0.05606596, -0.46260933, 0.16175873]]) |

Run example 112: svm\_with\_sgd\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /svm\_with\_sgd\_example .py |

Operation results:

|  |
| --- |
| Training Error = 0.38198757763975155 |

Run example 113: mllib\_tf\_idf\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /tf\_idf\_example .py |

Operation results:

|  |
| --- |
| tfidf:  (1048576,[136293],[3.758288905486104])  (1048576,[904617],[3.758288905486104])  (1048576,[604736],[3.758288905486104])  (1048576,[304703],[3.758288905486104])  (1048576,[905950],[3.758288905486104])  (1048576,[37922],[3.758288905486104])  tfidfIgnore:  (1048576,[136293],[0.0])  (1048576,[904617],[0.0])  (1048576,[604736],[0.0])  (1048576,[304703],[0.0])  (1048576,[905950],[0.0])  (1048576,[37922],[0.0]) |

Run Example 114: word2vec

|  |
| --- |
| python ./examples/src/main/python/ml lib /word2vec .py |

Operation results:

|  |
| --- |
| bin/spark-submit --driver-memory 4g examples/src/main/python/mllib/word2vec.py text8\_lines  Argument for file not provided |

This use case requires input data data/mllib/sample\_lda\_data.txt

Run Example 114: word2vec

|  |
| --- |
| python ./examples/src/main/python/ml lib /word2vec .py da ta/mllib/sample\_lda\_data.txt |

Operation results:

|  |
| --- |
| Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/examples/src/main/python/mllib/word2vec.py", line 45, in <module>  synonyms = model.findSynonyms('china', 40)  ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^  File "/home/openeuler/zhuchenhao/venv\_3.11\_pyspark/lib/python3.11/site-packages/pyspark/mllib/feature.py", line 841, in findSynonyms  words, similarity = self.call("findSynonyms", word, num)  ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^  File "/home/openeuler/zhuchenhao/venv\_3.11\_pyspark/lib/python3.11/site-packages/py4j/protocol.py", line 326, in get\_return\_value  raise Py4JJavaError(  py4j.protocol.Py4JJavaError: An error occurred while calling o33.findSynonyms.  : java.lang.IllegalStateException: china not in vocabulary |

By comparing the two use cases word2vec.py and word2vec\_example.py, we found that there is no difference in the algorithm code content between the two. word2vec.py should be an obsolete test case that has not been deleted. The code reminds us to run bin/spark-submit --driver-memory 4g examples/src/main/python/mllib/word2vec.py text8\_lines . However, the Spark library no longer has the test8\_lines.txt file.

Run example 115: mllib\_word2vec\_example

|  |
| --- |
| python ./examples/src/main/python/ml lib /word2vec\_example .py |

Operation results:

|  |
| --- |
| 0: 0.13997915387153625  4: 0.08936721086502075  3: 0.07495273649692535  9: 0.05684074014425278  2: 0.05529635399580002 |

1. Python test case 3: Test cases in the pyspark/.../ml / folder

Run case 116 : test\_algorithms.py

|  |
| --- |
| # Note that it is best to enter the spark-3.5.1 folder to run the test case  python ./python/pyspark/ml/tests/binary\_classification\_metrics\_example .py​​ |

Operation results:

|  |
| --- |
| …  Ran 17 tests in 837.962s |

Run test case 117: test\_base

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_base .py​​ |

Operation results:

|  |
| --- |
| Ran 5 tests in 154.148s |

Run test case 118: test\_dl\_util

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_dl\_util .py​​ |

Operation results:

|  |
| --- |
| Ran 8 tests in 0.020s |

Run case 119: test\_evaluation

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_base .py​​ |

Operation results:

|  |
| --- |
| Ran 3 tests in 167.260s |

Run test case 120: test\_feature

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_base .py​​ |

Operation results:

|  |
| --- |
| Ran 13 tests in 376.430s |

Run case 121: test\_functions

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_functions .py​​ |

Operation results:

|  |
| --- |
| #RISC-V  ValueError: Multiple input columns found, but model expected a single input, use `array` to combine columns into tensors  Ran 11 tests in 385.724s  FAILED (failures=1, errors=1)  # ARM  Ran 11 tests in 38.781s  OK |

Run test case 122: test\_image

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_image .py​​ |

Operation results:

|  |
| --- |
| Ran 1 test in 206.845s |

Run test case 123: test\_linalg

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_linalg .py​​ |

Operation results:

|  |
| --- |
| Ran 22 tests in 587.584s |

Run test case 124: test\_model\_cache

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_model\_cache .py​​ |

Operation results:

|  |
| --- |
| Ran 1 test in 105.005s |

Run case 125: test\_param

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_param .py​​ |

Operation results:

|  |
| --- |
| Ran 27 tests in 293.285s |

Run test case 126: test\_persistence

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_persistence.py​​​ |

Operation results:

|  |
| --- |
| Ran 17 tests in 1389.043s |

Run test case 127: test\_pipeline

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_pipeline.py​​​ |

Operation results:

|  |
| --- |
| Ran 2 tests in 80.450s |

Run case 128: test\_stat

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_stat .py​​ |

Operation results:

|  |
| --- |
| Ran 1 test in 188.107s |

Run case 129: test\_training\_summary

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_base .py​​ |

Operation results:

|  |
| --- |
| #RISC-V  FAIL: test\_gaussian\_mixture\_summary (pyspark.ml.tests.test\_training\_summary.TrainingSummaryTest.test\_gaussian\_mixture\_summary)  ----------------------------------------------------------------------  Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/tests/test\_training\_summary.py", line 442, in test\_gaussian\_mixture\_summary  self.assertEqual(s.numIter, 3)  AssertionError: 6 != 3  ----------------------------------------------------------------------  Ran 12 tests in 608.769s  FAILED (failures=1)  1.0  # ARM  Ran 12 tests in 41.111s  OK  1.0 |

This use case encountered a calculation error on the RISC-V server, but it went through all the calculations so it can be compared with the ARM use case performance.

Run test case 130: test\_util

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_util .py​​ |

Operation results:

|  |
| --- |
| Ran 1 test in 290.661s |

Run test case 131: test\_wrapper

|  |
| --- |
| python ./python/pyspark/ml/tests/test\_wrapper .py​​ |

Operation results:

|  |
| --- |
| Ran 2 tests in 117.764s |

1. Python test case 3: Test case in pyspark/.../ml lib/ folder

Run case 132: mllib\_test\_algorithms

|  |
| --- |
| python ./python/pyspark/m l l ib / tests /test\_algorithms .py |

Operation results:

|  |
| --- |
| Ran 11 tests in 861.439s |

Run case 133: mllib\_test\_feature

|  |
| --- |
| python ./python/pyspark/m l l ib / tests/ test\_feature .py |

Operation results:

|  |
| --- |
| Ran 9 tests in 377.540s |

Run case 134: mllib\_test\_linalg

|  |
| --- |
| python ./python/pyspark/m l l ib / tests/ mllib\_test\_linalg .py |

Operation results:

|  |
| --- |
| Ran 34 tests in 267.850s |

Run case 135: mllib\_test\_stat

|  |
| --- |
| python ./python/pyspark/m l l ib / tests/ mllib\_test\_stat .py |

Operation results:

|  |
| --- |
| Ran 7 tests in 240.068s |

Run case 136: mllib\_test\_streaming\_algorithms

|  |
| --- |
| # Note that it is best to enter the spark-3.5.1 folder to run the test case  python ./python/pyspark/m l l ib / tests/ test\_streaming\_algorithms .py |

Operation results:

|  |
| --- |
| #RISC-V  ASSERTIONERROR: TEST FAILED DUE TO TIMEOUT AFTER 180 SEC, WITH LAST CONDITION RETURNING: LATEST ERRORS: 0.67, 0.71, 0.78, 0.7, 0.75, 0.74, 0.73, 0.69, 0.62, 0.71, 0.69, 0.75  ----------------------------------------------------------------------  RAN 13 TESTS IN 994.816S  FAILED(FAILURES=8, SKIPPED=1)  # ARM  Ran 13 tests in 275.645s  OK (skipped=1) |

Run case 137: mllib\_test\_util

|  |
| --- |
| python ./python/pyspark/m l l ib /tests / test\_util .py |

Operation results:

|  |
| --- |
| Ran 6 tests in 105.837s |

1. Python test case 3: Test case in the pyspark/.../ml/connect folder

Run test case 138: test\_connect\_classification

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_connect\_classification.py​​​ |

Operation results:

|  |
| --- |
| Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/tests/connect/test\_connect\_classification.py", line 40, in setUp  .getOrCreate()  ^^^^^^^^^^^^^  Ran 3 tests in 39.478s  FAILED (errors=3) |

Run test case 139: test\_connect\_evaluation

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_connect\_evaluation.py​​​ |

Operation results:

|  |
| --- |
| Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/tests/connect/test\_connect\_evaluation.py", line 32, in setUp  self.spark = SparkSession.builder.remote("local[2]").getOrCreate()  ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^Ran 3 tests in 0.016s  OK (skipped=3) |

Run test case 140: test\_connect\_feature

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_connect\_feature .py​​ |

Operation results:

|  |
| --- |
| Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/tests/connect/test\_connect\_feature.py", line 25, in setUp  self.spark = SparkSession.builder.remote("local[2]").getOrCreate()  ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^  Ran 2 tests in 60.858s  FAILED (errors=2) |

Run case 141: test\_connect\_function

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_connect\_function .py​​ |

Operation results:

|  |
| --- |
| Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/tests/connect/test\_connect\_function.py", line 42, in setUpClass  super(SparkConnectMLFunctionTests, cls).setUpClass()  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/testing/connectutils.py", line 189, in setUpClass  .getOrCreate()  ^^^^^^^^^^^^^  Ran 0 tests in 48.015s  FAILED (errors=1) |

Run test case 142: test\_connect\_pipeline

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_connect\_pipeline .py​​ |

Operation results:

|  |
| --- |
| Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/tests/connect/test\_connect\_pipeline.py", line 29, in setUp  .getOrCreate()  ^^^^^^^^^^^^^  Ran 2 tests in 85.694s  FAILED (errors=2) |

Run test case 143: test\_connect\_summarizer

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_connect\_pipeline .py​​ |

Operation results:

|  |
| --- |
| Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/tests/connect/test\_connect\_summarizer.py", line 25, in setUp  self.spark = SparkSession.builder.remote("local[2]").getOrCreate()  ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^  Ran 1 test in 43.435s  FAILED (errors=1) |

Run test case 144: test\_connect\_tuning

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_connect\_tuning .py​​ |

Operation results:

|  |
| --- |
| Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/tests/connect/test\_connect\_tuning.py", line 29, in setUp  .getOrCreate()  ^^^^^^^^^^^^^  Ran 6 tests in 53.558s  FAILED (errors=6) |

Run test case 145: test\_legacy\_mode\_classification

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_legacy\_mode\_classification.py​​​ |

Operation results:

|  |
| --- |
| Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/tests/connect/test\_legacy\_mode\_classification.py", line 22, in <module>  from pyspark.ml.connect.classification import (  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/connect/classification.py", line 46, in <module>  import torch  ModuleNotFoundError: No module named 'torch' |

The Connect use case requires additional dependencies, so you need to install pytorch, scikit-learn, grpcio protobuf , grpcio-status. torcheval . After installation, run the test case again:

Operation results:

|  |
| --- |
| # RISC-V output  Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/tests/connect/test\_legacy\_mode\_classification.py", line 169, in test\_save\_load  model = estimator.fit(training\_dataset)  ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^  Ran 3 tests in 300.984s  FAILED (errors=3)  # ARM output  Ran 3 tests in 79.780s  OK |

Run test case 146: test\_legacy\_mode\_evaluation

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_legacy\_mode\_evaluation.py​​​ |

Operation results:

|  |
| --- |
| Ran 3 tests in 235.444s  OK |

Run test case 147: test\_legacy\_mode\_feature

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_legacy\_mode\_feature .py​​ |

Operation results:

|  |
| --- |
| Ran 2 tests in 229.922s  OK |

Run test case 148: test\_legacy\_mode\_pipeline

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_legacy\_mode\_pipeline .py​​ |

Operation results:

|  |
| --- |
| # RISC-V output  Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/tests/connect/test\_legacy\_mode\_pipeline.py", line 70, in test\_pipeline  model = pipeline.fit(train\_dataset)  ^^^^^^^^^^^^^^^^^^^^^^^^^^^  Ran 2 tests in 152.792s  FAILED (errors=1)  # ARM output  Ran 2 tests in 56.862s  OK |

Run test case 149: test\_legacy\_mode\_summarizer

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_legacy\_mode\_summarizer .py​​ |

Operation results:

|  |
| --- |
| Ran 1 test in 117.087s  OK |

Run case 150: test\_legacy\_mode\_tuning

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_legacy\_mode\_tuning .py​​ |

Operation results:

|  |
| --- |
| Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/tests/connect/test\_legacy\_mode\_tuning.py", line 272, in test\_crossvalidator\_with\_fold\_col  cv.fit(train\_dataset)  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/connect/base.py", line 106, in fit  return self.\_fit(dataset)  ^^^^^^^^^^^^^^^^^^  Ran 6 tests in 451.079s  FAILED (errors=2)  # ARM output  Ran 6 tests in 277.443s  OK |

Run case 151: test\_parity\_torch\_data\_loader

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_parity\_torch\_data\_loader .py​​ |

Operation results:

|  |
| --- |
| Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/torch/tests/test\_data\_loader.py", line 41, in setUp  .getOrCreate()  ^^^^^^^^^^^^^  Ran 2 tests in 48.482s  FAILED (errors=2) |

Run use case 152: test\_parity\_torch\_distributor

|  |
| --- |
| python ./python/pyspark/ml/tests/connect/test\_parity\_torch\_distributor .py​​ |

Operation results:

|  |
| --- |
| Traceback (most recent call last):  File "/home/openeuler/zhuchenhao/spark-3.5.1/python/pyspark/ml/tests/connect/test\_parity\_torch\_distributor.py", line 67, in setUpClass  cls.spark = builder.remote("local-cluster[2,2,512]").getOrCreate()  ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^  Ran 2 tests in 52.759s  FAILED (errors=4) |

1. Python test case 3: Test cases in the pyspark/.../ml/ tuning folder

Run test case 153: test\_cv\_io\_basic

|  |
| --- |
| python ./python/pyspark/ml/tests/tuning/test\_cv\_io\_basic .py​​ |

Operation results:

|  |
| --- |
| Ran 2 tests in 844.387s |

Run test case 154: test\_cv\_io\_nested

|  |
| --- |
| python ./python/pyspark/ml/tests/tuning/test\_cv\_io\_nested .py​​ |

Operation results:

|  |
| --- |
| Ran 1 test in 683.965s |

Run test case 155: test\_cv\_io\_pipeline

|  |
| --- |
| python ./python/pyspark/ml/tests/tuning/test\_cv\_io\_pipeline .py​​ |

Operation results:

|  |
| --- |
| Ran 1 test in 1480.401s |

Run case 156: test\_tuning

|  |
| --- |
| python ./python/pyspark/ml/tests/tuning/test\_tuning .py​​ |

Operation results:

|  |
| --- |
| Ran 15 tests in 943.030s  OK  0.03676847843514508 |

Run test case 157: test\_tvs\_io\_basic

|  |
| --- |
| python ./python/pyspark/ml/tests/tuning/test\_tvs\_io\_basic .py​​ |

Operation results:

|  |
| --- |
| Ran 2 tests in 486.055s |

Run test case 158: test\_tvs\_io\_nested

|  |
| --- |
| python ./python/pyspark/ml/tests/tuning/test\_tvs\_io\_nested .py​​ |

Operation results:

|  |
| --- |
| Ran 1 test in 533.386s |

Run test case 159: test\_tvs\_io\_pipeline

|  |
| --- |
| python ./python/pyspark/ml/tests/tuning/test\_tvs\_io\_pipeline .py​​ |

Operation results:

|  |
| --- |
| Ran 1 test in 1118.589s |

1. Java test cases: test cases in the ml/ folder

Run Example 160: JavaAFTSurvivalRegressionExample

(equivalent to Python's aft\_survival\_regression )

|  |
| --- |
| # Note that it is best to enter the spark-3.5.1 folder to run the test case  ./bin/spark-submit --class org.apache.spark.examples.ml.JavaAFTSurvivalRegressionExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml.JavaAFTSurvivalRegressionExample.  # ARM output  Coefficients: [-0.4963068060199873,0.19844393975928323]  Intercept: 2.638090563156023  Scale: 1.5472326865488453  +-----+------+--------------+------------------+--------------------------------------+  |label|censor|features |prediction |quantiles |  +-----+------+--------------+------------------+--------------------------------------+  |1.218|1.0 |[1.56,-0.605] |5.718996553029903 |[1.16032959510291,4.995471733719649] |  |2.949|0.0 |[0.346,2.158] |18.076458028588913|[3.66754010615639,15.789559285491208] |  |3.627|0.0 |[1.38,0.231] |7.381875365763506 |[1.4977117707333802,6.447975512763029]|  |0.273|1.0 |[0.52,1.151] |13.577581299077902|[2.754761130759774,11.859846908963423]|  |4.199|0.0 |[0.795,-0.226]|9.013093216625737 |[1.828670240609156,7.872823838856886] |  +-----+------+--------------+------------------+--------------------------------------+ |

Run Example 161: JavaALSExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaALSExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml.JavaALSExample.  # ARM output  Root-mean-square error = 1.9227839872857166  +------+--------------------+  |userId| recommendations|  +------+--------------------+  | 20|[{22, 4.9981427},...|  | 10|[{2, 3.5062227}, ...|  | 0|[{92, 3.6381528},...|  | 1|[{62, 3.8919413},...|  | 21|[{53, 5.0316496},...|  | 11|[{18, 5.2035656},...|  | 12|[{17, 4.9443045},...|  | 22|[{75, 5.1680317},...|  | 2|[{53, 7.1292467},...|  | 13|[{93, 3.9177742},...|  | 3|[{51, 4.7208066},...|  | 23|[{46, 5.8863363},...|  | 4|[{23, 5.574499}, ...|  | 24|[{69, 4.922538}, ...|  | 14|[{52, 4.9217186},...|  | 5|[{55, 4.4586773},...|  | 15|[{1, 3.7592566}, ...|  | 25|[{39, 4.0591235},...|  | 26|[{75, 5.5346346},...|  | 6|[{25, 5.2203145},...|  +------+--------------------+  only showing top 20 rows |

Run Example 162: JavaBinarizerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaBinarizerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +---+-------+-----------------+  | id|feature|binarized\_feature|  +---+-------+-----------------+  | 0| 0.1| 0.0|  | 1| 0.8| 1.0|  | 2| 0.2| 0.0|  +---+-------+-----------------+ |

Run Example 163: JavaBisectingKMeansExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaBisectingKMeansExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Silhouette with squared euclidean distance = 0.9997530305375207  Cluster Centers:  [0.1,0.1,0.1]  [9.1,9.1,9.1] |

Run Example 164: JavaBucketedRandomProjectionLSHExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaBucketedRandomProjectionLSHExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  The hashed dataset where hashed values are stored in the column 'hashes':  +---+-----------+--------------------+  | id| features| hashes|  +---+-----------+--------------------+  | 0| [1.0,1.0]|[[0.0], [0.0], [-...|  | 1| [1.0,-1.0]|[[-1.0], [-1.0], ...|  | 2|[-1.0,-1.0]|[[-1.0], [-1.0], ...|  | 3| [-1.0,1.0]|[[0.0], [0.0], [-...|  +---+-----------+--------------------+  Approximately joining dfA and dfB on distance smaller than 1.5:  +---+---+-----------------+  |idA|idB|EuclideanDistance|  +---+---+-----------------+  | 1| 4| 1.0|  | 0| 6| 1.0|  | 1| 7| 1.0|  | 3| 5| 1.0|  | 0| 4| 1.0|  | 3| 6| 1.0|  | 2| 7| 1.0|  | 2| 5| 1.0|  +---+---+-----------------+  Approximately searching dfA for 2 nearest neighbors of the key:  +---+----------+--------------------+-------+  | id| features| hashes|distCol|  +---+----------+--------------------+-------+  | 0| [1.0,1.0]|[[0.0], [0.0], [-...| 1.0|  | 1|[1.0,-1.0]|[[-1.0], [-1.0], ...| 1.0|  +---+----------+--------------------+-------+ |

Run Example 165: JavaBucketizerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaBucketizerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Bucketizer output with 4 buckets  +--------+----------------+  |features|bucketedFeatures|  +--------+----------------+  |-999.9| 0.0|  | -0.5| 1.0|  | -0.3| 1.0|  | 0.0| 2.0|  | 0.2| 2.0|  | 999.9| 3.0|  +--------+----------------+  Bucketizer output with [4, 4] buckets for each input column  +---------+---------+-----------------+-----------------+  |features1|features2|bucketedFeatures1|bucketedFeatures2|  +---------+---------+-----------------+-----------------+  | -999.9| -999.9| 0.0| 0.0|  | -0.5| -0.2| 1.0| 1.0|  | -0.3| -0.1| 1.0| 1.0|  | 0.0| 0.0| 2.0| 2.0|  | 0.2| 0.4| 2.0| 3.0|  | 999.9| 999.9| 3.0| 3.0|  +---------+---------+-----------------+-----------------+ |

Run Example 166: JavaChiSqSelectorExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaChiSqSelectorExample--master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  ChiSqSelector output with top 1 features selected  +---+------------------+-------+----------------+  | id| features|clicked|selectedFeatures|  +---+------------------+-------+----------------+  | 7|[0.0,0.0,18.0,1.0]| 1.0| [18.0]|  | 8|[0.0,1.0,12.0,0.0]| 0.0| [12.0]|  | 9|[1.0,0.0,15.0,0.1]| 0.0| [15.0]|  +---+------------------+-------+----------------+ |

Run Example 167: JavaChiSquareTestExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaChiSquareTestExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  pValues: [0.6872892787909721,0.6822703303362126]  degreesOfFreedom: [2, 3]  statistics: [0.75,1.5] |

Run Example 168: JavaCorrelationExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaCorrelationExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Pearson correlation matrix:  1.0 0.05564148840746571 NaN 0.40047142031681365  0.05564148840746571 1.0 NaN 0.9135958615342521  NaN NaN 1.0 NaN  0.40047142031681365 0.9135958615342521 NaN 1.0  Spearman correlation matrix:  1.0 0.10540925533894598 NaN 0.4  0.10540925533894598 1.0 NaN 0.9486832980505139  NaN NaN 1.0 NaN  0.4 0.9486832980505139 NaN 1.0 |

Run Example 169: JavaCountVectorizerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaCountVectorizerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +---------------+-------------------------+  |text |feature |  +---------------+-------------------------+  |[a, b, c] |(3,[0,1,2],[1.0,1.0,1.0])|  |[a, b, b, c, a]|(3,[0,1,2],[2.0,2.0,1.0])|  +---------------+-------------------------+ |

Run use case 170: JavaDCTExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaDCTExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +----------------------------------------------------------------+  |featuresDCT |  +----------------------------------------------------------------+  |[1.0,-1.1480502970952693,2.0000000000000004,-2.7716385975338604]|  |[-1.0,3.378492794482933,-7.000000000000001,2.9301512653149677] |  |[4.0,9.304453421915744,11.000000000000002,1.5579302036357163] |  +----------------------------------------------------------------+ |

Run Example 171: JavaDecisionTreeClassificationExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaDecisionTreeClassificationExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +--------------+-----+--------------------+  |predictedLabel|label| features|  +--------------+-----+--------------------+  | 0.0| 0.0|(692,[95,96,97,12...|  | 0.0| 0.0|(692,[122,123,124...|  | 0.0| 0.0|(692,[124,125,126...|  | 0.0| 0.0|(692,[125,126,127...|  | 0.0| 0.0|(692,[126,127,128...|  +--------------+-----+--------------------+  only showing top 5 rows  Test Error = 0.0  Learned classification tree model:  DecisionTreeClassificationModel: uid=dtc\_d50006c90f8e, depth=2, numNodes=5, numClasses=2, numFeatures=692  If (feature 434 <= 70.5)  If (feature 99 in {2.0})  Predict: 0.0  Else (feature 99 not in {2.0})  Predict: 1.0  Else (feature 434 > 70.5)  Predict: 0.0 |

Run Example 172: JavaDecisionTreeRegressionExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaDecisionTreeRegressionExample--master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +-----+--------------------+  |label| features|  +-----+--------------------+  | 0.0|(692,[100,101,102...|  | 0.0|(692,[121,122,123...|  | 0.0|(692,[122,123,124...|  | 0.0|(692,[123,124,125...|  | 0.0|(692,[124,125,126...|  +-----+--------------------+  only showing top 5 rows  Root Mean Squared Error (RMSE) on test data = 0.0  Learned regression tree model:  DecisionTreeRegressionModel: uid=dtr\_60709e2a2e1d, depth=2, numNodes=5, numFeatures=692  If (feature 434 <= 70.5)  If (feature 99 in {0.0,3.0})  Predict: 0.0  Else (feature 99 not in {0.0,3.0})  Predict: 1.0  Else (feature 434 > 70.5)  Predict: 1.0 |

Run Example 173: JavaDocument

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaDocument --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Exception in thread "main" java.lang.NoSuchMethodException: org.apache.spark.examples.ml.JavaDocument.main([Ljava.lang.String;) |

Run Example 174: JavaElementwiseProductExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaElementwiseProductExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +---+-------------+-----------------+  | id| vector|transformedVector|  +---+-------------+-----------------+  | a|[1.0,2.0,3.0]| [0.0,2.0,6.0]|  | b|[4.0,5.0,6.0]| [0.0,5.0,12.0]|  +---+-------------+-----------------+ |

Run Example 175: JavaEstimatorTransformerParamExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaEstimatorTransformerParamExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  LogisticRegression parameters:  aggregationDepth: suggested depth for treeAggregate (>= 2) (default: 2) |

Run Example 176: JavaFeatureHasherExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaFeatureHasherExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +----+-----+---------+------+--------------------------------------------------------+  |real|bool |stringNum|string|features |  +----+-----+---------+------+--------------------------------------------------------+  |2.2 |true |1 |foo |(262144,[174475,247670,257907,262126],[2.2,1.0,1.0,1.0])|  |3.3 |false|2 |bar |(262144,[70644,89673,173866,174475],[1.0,1.0,1.0,3.3]) |  |4.4 |false|3 |baz |(262144,[22406,70644,174475,187923],[1.0,1.0,4.4,1.0]) |  |5.5 |false|4 |foo |(262144,[70644,101499,174475,257907],[1.0,1.0,5.5,1.0]) |  +----+-----+---------+------+--------------------------------------------------------+ |

Run Example 177: JavaFMClassifierExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml . JavaFMClassifierExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +--------------+-----+--------------------+  |predictedLabel|label| features|  +--------------+-----+--------------------+  | 0.0| 0.0|(692,[122,123,124...|  | 0.0| 0.0|(692,[124,125,126...|  | 0.0| 0.0|(692,[124,125,126...|  | 0.0| 0.0|(692,[126,127,128...|  | 0.0| 0.0|(692,[126,127,128...|  +--------------+-----+--------------------+  only showing top 5 rows |

Run Example 178: JavaFMRegressorExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaFMRegressorExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +--------------------+-----+--------------------+  | prediction|label| features|  +--------------------+-----+--------------------+  | 0.08907253082444606| 0.0|(692,[122,123,148...|  | 0.19391784102721332| 0.0|(692,[124,125,126...|  | 0.2532968736395909| 0.0|(692,[125,126,127...|  |-0.10145448143775096| 0.0|(692,[126,127,128...|  | 0.07264440772941294| 0.0|(692,[126,127,128...|  +--------------------+-----+--------------------+  only showing top 5 rows |

Run Example 179: JavaFPGrowthExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaFPGrowthExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +---------+----+  |items|freq|  +---------+----+  | [5]| 2|  | [5, 1]| 2|  |[5, 1, 2]| 2|  | [5, 2]| 2|  | [1]| 3|  | [1, 2]| 3|  | [2]| 3|  +---------+----+  +----------+----------+------------------+----+------------------+  |antecedent|consequent| confidence|lift| support|  +----------+----------+------------------+----+------------------+  | [1, 2]| [5]|0.6666666666666666| 1.0|0.6666666666666666|  | [5, 1]| [2]| 1.0| 1.0|0.6666666666666666|  | [2]| [5]|0.6666666666666666| 1.0|0.6666666666666666|  | [2]| [1]| 1.0| 1.0| 1.0|  | [5]| [1]| 1.0| 1.0|0.6666666666666666|  | [5]| [2]| 1.0| 1.0|0.6666666666666666|  | [1]| [5]|0.6666666666666666| 1.0|0.6666666666666666|  | [1]| [2]| 1.0| 1.0| 1.0|  | [5, 2]| [1]| 1.0| 1.0|0.6666666666666666|  +----------+----------+------------------+----+------------------+  +------------+----------+  |items|prediction|  +------------+----------+  | [1, 2, 5]| []|  |[1, 2, 3, 5]| []|  | [1, 2]| [5]|  +------------+----------+ |

Run Example 180: JavaGaussianMixtureExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaGaussianMixtureExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Gaussian 0:  weight=0.500000  mu=[0.10000000000001552,0.10000000000001552,0.10000000000001552]  sigma=  0.0066666666666806454 0.0066666666666806454 0.0066666666666806454  0.0066666666666806454 0.0066666666666806454 0.0066666666666806454  0.0066666666666806454 0.0066666666666806454 0.0066666666666806454  Gaussian 1:  weight=0.500000  mu=[9.099999999999984,9.099999999999984,9.099999999999984]  sigma=  0.0066666666666808728 0.0066666666666808728 0.0066666666666808728  0.0066666666666808728 0.0066666666666808728 0.0066666666666808728  0.0066666666666808728 0.0066666666666808728 0.0066666666666808728 |

Run Example 181: JavaGeneralizedLinearRegressionExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaGeneralizedLinearRegressionExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Gaussian 0:  weight=0.500000  mu=[0.10000000000001552,0.10000000000001552,0.10000000000001552]  sigma=  0.0066666666666806454 0.0066666666666806454 0.0066666666666806454  0.0066666666666806454 0.0066666666666806454 0.0066666666666806454  0.0066666666666806454 0.0066666666666806454 0.0066666666666806454  Gaussian 1:  weight=0.500000  mu=[9.099999999999984,9.099999999999984,9.099999999999984]  sigma=  0.0066666666666808728 0.0066666666666808728 0.0066666666666808728  0.0066666666666808728 0.0066666666666808728 0.0066666666666808728  0.0066666666666808728 0.0066666666666808728 0.0066666666666808728 |

Run Example 182: JavaGradientBoostedTreeClassifierExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaGradientBoostedTreeClassifierExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +--------------+-----+--------------------+  |predictedLabel|label| features|  +--------------+-----+--------------------+  | 0.0| 0.0|(692,[95,96,97,12...|  | 0.0| 0.0|(692,[100,101,102...|  | 0.0| 0.0|(692,[121,122,123...|  | 0.0| 0.0|(692,[122,123,148...|  | 0.0| 0.0|(692,[123,124,125...|  +--------------+-----+--------------------+  only showing top 5 rows |

Run Example 183: JavaGradientBoostedTreeRegressorExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaGradientBoostedTreeRegressorExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +----------+-----+--------------------+  |prediction|label| features|  +----------+-----+--------------------+  | 0.0| 0.0|(692,[100,101,102...|  | 0.0| 0.0|(692,[121,122,123...|  | 0.0| 0.0|(692,[123,124,125...|  | 0.0| 0.0|(692,[124,125,126...|  | 0.0| 0.0|(692,[124,125,126...|  +----------+-----+--------------------+  only showing top 5 rows |

Run Example 184: JavaImputerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaImputerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +---+---+-----+-----+  | a| b|out\_a|out\_b|  +---+---+-----+-----+  |1.0|NaN| 1.0| 4.0|  |2.0|NaN| 2.0| 4.0|  |NaN|3.0| 3.0| 3.0|  |4.0|4.0| 4.0| 4.0|  |5.0|5.0| 5.0| 5.0|  +---+---+-----+-----+ |

Run Example 185: JavaIndexToStringExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaIndexToStringExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Transformed string column 'category' to indexed column 'categoryIndex'  +---+--------+-------------+  | id|category|categoryIndex|  +---+--------+-------------+  | 0| a| 0.0|  | 1| b| 2.0|  | 2| c| 1.0|  | 3| a| 0.0|  | 4| a| 0.0|  | 5| c| 1.0|  +---+--------+-------------+  StringIndexer will store labels in output column metadata: {"vals":["a","c","b"],"type":"nominal","name":"categoryIndex"} |

Run Use Case 186: JavaInteractionExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaInteractionExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +---+--------------+--------------+------------------------------------------------------+  |id1|vec1 |vec2 |interactedCol |  +---+--------------+--------------+------------------------------------------------------+  |1 |[1.0,2.0,3.0] |[8.0,4.0,5.0] |[8.0,4.0,5.0,16.0,8.0,10.0,24.0,12.0,15.0] |  |2 |[4.0,3.0,8.0] |[7.0,9.0,8.0] |[56.0,72.0,64.0,42.0,54.0,48.0,112.0,144.0,128.0] |  |3 |[6.0,1.0,9.0] |[2.0,3.0,6.0] |[36.0,54.0,108.0,6.0,9.0,18.0,54.0,81.0,162.0] |  |4 |[10.0,8.0,6.0]|[9.0,4.0,5.0] |[360.0,160.0,200.0,288.0,128.0,160.0,216.0,96.0,120.0]|  |5 |[9.0,2.0,7.0] |[10.0,7.0,3.0]|[450.0,315.0,135.0,100.0,70.0,30.0,350.0,245.0,105.0] |  |6 |[1.0,1.0,4.0] |[2.0,8.0,4.0] |[12.0,48.0,24.0,12.0,48.0,24.0,48.0,192.0,96.0] |  +---+--------------+--------------+------------------------------------------------------+ |

Run Example 187: JavaIsotonicRegressionExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaIsotonicRegressionExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Boundaries in increasing order: [0.01,0.17,0.18,0.27,0.28,0.29,0.3,0.31,0.34,0.35,0.36,0.41,0.42,0.71,0.72,0.74 ,0.75,0.76,0.77,0.78,0.79,0.8,0.81,0.82,0.83,0.84,0.85,0.86,0.87,0.88,0.89,1.0] |

Run Example 188: JavaKMeansExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaKMeansExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Silhouette with squared euclidean distance = 0.9997530305375207  Cluster Centers:  [9.1,9.1,9.1]  [0.1,0.1,0.1] |

Run Example 189: JavaLabeledDocument

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaLabeledDocument --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Exception in thread "main" java.lang.NoSuchMethodException: org.apache.spark.examples.ml.JavaLabeledDocument.main([Ljava.lang.String;) |

Run Example 190: JavaLDAExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaLDAExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  The lower bound on the log likelihood of the entire corpus: -788.952861952069  The upper bound on perplexity: 3.0344340844310347  The topics described by their top-weighted terms:  +-----+-----------+---------------------------------------------------------------+  |topic|termIndices|termWeights |  +-----+-----------+---------------------------------------------------------------+  |0 |[2, 5, 7] |[0.1059658039810754, 0.1056051818283619, 0.10421666164252749] |  |1 |[1, 6, 2] |[0.10177948132705314, 0.09812200626123077, 0.09628914045922496]|  |2 |[1, 9, 4] |[0.10587343931538634, 0.0974641589886441, 0.0965079506935491] |  |3 |[3, 0, 5] |[0.1743854287956877, 0.13136091853187473, 0.12944901538860193] |  |4 |[9, 6, 4] |[0.10444192230873739, 0.10406373129411564, 0.10097458055419123]|  |5 |[10, 6, 1] |[0.18472780200432165, 0.18465951178879786, 0.14330925941690284]|  |6 |[3, 7, 4] |[0.11621744682267901, 0.0989646711684413, 0.09790791643962576] |  |7 |[4, 0, 2] |[0.10844106087588981, 0.10326272852181545, 0.1002885773336093] |  |8 |[0, 7, 8] |[0.10995540954614377, 0.09914316836644671, 0.09806208481379909]|  |9 |[9, 6, 8] |[0.10099428301371008, 0.10007218022977753, 0.09764791948287375]|  +-----+-----------+---------------------------------------------------------------+ |

Run Example 191: JavaLinearRegressionWithElasticNetExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaLinearRegressionWithElasticNetExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Coefficients: [0.0,0.3229251667740594,-0.3438548034562219,1.915601702345841,0.05288058680386255, 0.765962720459771,0.0,-0.15105392669186676,-0.21587930360904645,0.2202536918881343] Intercept: 0.15989368442397356  numIterations: 6  objectiveHistory: [0.49999999999999994,0.4967620357443381,0.49363616643404634,0.493635 1537897608,0.4936351214177871,0.49363512062528014,0.4936351206216114] |

Run Example 192: JavaLinearSVCExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaLinearSVCExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Coefficients: [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0. 0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0 .0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0, |

Run Example 193: JavaLogisticRegressionSummaryExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaLogisticRegressionSummaryExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  0.6833149135741672  0.6661906127558116  0.6207433672479603  0.613154125312387  0.6059149689952392  0.5923656241678247  0.5898233082838024  0.5868012627420287  0.5844432058719141  0.5830790068041742  0.5807015754032356 |

Run Example 194: JavaLogisticRegressionWithElasticNetExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaLogisticRegressionWithElasticNetExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Coefficients: (692,[272,300,323,350,351,378,379,405,406,407,428,433,434,435,455,456,461,462,483,484,489,490 ,496,511,512,517,539,540,568],[-7.520689871384053E-5,-8.11577314684692E-5,3.81469277184646E-5 ,3.776490540424339E-4,3.405148366194403E-4,5.51445515734312E-4,4.085386116096919E-4,4.1974673 32749458E-4,8.119171358670047E-4,5.027708372668761E-4,-2.3929260406600262E-5,5.745048020902315 E-4,9.037546426803657E-4,7.81822970024382E-5,-2.1787551952911182E-5,-3.402165821789468E-5,4.9 66517360637647E-4,8.190557828370378E-4,-8.017982139522552E-5,-2.7431694037834526E-5,4.8108322 26238997E-4,4.840801762677886E-4,-8.926472920008605E-6,-3.414881233042729E-4,-8.9505925741213 12E-5,4.864546911689224E-4,-8.478698005186017E-5,-4.2347832158317754E-4,-7.29653577763117E-5]) Intercept: -0.5991460286401454 |

Run Example 195: JavaMaxAbsScalerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaMaxAbsScalerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +--------------+--------------------+  | features| scaledFeatures|  +--------------+--------------------+  |[1.0,0.1,-8.0]|[0.25,0.010000000...|  |[2.0,1.0,-4.0]| [0.5,0.1,-0.5]|  |[4.0,10.0,8.0]| [1.0,1.0,1.0]|  +--------------+--------------------+ |

Run Example 196: JavaMinHashLSHExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaMinHashLSHExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  The hashed dataset where hashed values are stored in the column 'hashes':  +---+--------------------+--------------------+  | id| features| hashes|  +---+--------------------+--------------------+  | 0|(6,[0,1,2],[1.0,1...|[[2.25592966E8], ...|  | 1|(6,[2,3,4],[1.0,1...|[[2.25592966E8], ...|  | 2|(6,[0,2,4],[1.0,1...|[[2.25592966E8], ...|  +---+--------------------+--------------------+ |

Run Example 197: JavaMinMaxScalerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaMinMaxScalerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Features scaled to range: [0.0, 1.0]  +--------------+--------------+  | features|scaledFeatures|  +--------------+--------------+  |[1.0,0.1,-1.0]| (3,[],[])|  | [2.0,1.1,1.0]| [0.5,0.1,0.5]|  |[3.0,10.1,3.0]| [1.0,1.0,1.0]|  +--------------+--------------+ |

Run Example 198: JavaModelSelectionViaCrossValidationExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaModelSelectionViaCrossValidationExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  (4, spark ijk) --> prob=[0.2664764605192917,0.7335235394807083], prediction=1.0  (5, lmn) --> prob=[0.9203725758458563,0.0796274241541437], prediction=0.0  (6, mapreduce spark) --> prob=[0.44376360608061827,0.5562363939193817], prediction=1.0  (7, apache hadoop) --> prob=[0.8586524968002056,0.1413475031997944], prediction=0.0 |

Run Example 199: JavaModelSelectionViaTrainValidationSplitExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaModelSelectionViaTrainValidationSplitExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +--------------------+--------------------+--------------------+  | features| label| prediction|  +--------------------+--------------------+--------------------+  |(10,[0,1,2,3,4,5,...| -17.026492264209548| -1.6639541065509496|  |(10,[0,1,2,3,4,5,...| -16.71909683360509| -0.069169883774029|  |(10,[0,1,2,3,4,5,...| -15.375857723312297| 0.7290911157666173|  |(10,[0,1,2,3,4,5,...| -13.772441561702871| 3.8912558356758735|  |(10,[0,1,2,3,4,5,...| -13.039928064104615| 0.3591248102266543|  |(10,[0,1,2,3,4,5,...| -9.42898793151394| -3.975969158816923|  |(10,[0,1,2,3,4,5,...| -9.2679651250406|-0.05994784762268873|  |(10,[0,1,2,3,4,5,...| -9.173693798406978| -0.2877017140078841|  |(10,[0,1,2,3,4,5,...| -7.1500991588127265| 3.4142251414495983|  |(10,[0,1,2,3,4,5,...| -6.930603551528371| 0.0421398749842197|  |(10,[0,1,2,3,4,5,...| -6.456944198081549|-0.48310432084956806|  |(10,[0,1,2,3,4,5,...| -3.2843694575334834| -0.8029272819153814|  |(10,[0,1,2,3,4,5,...| -1.99891354174786| 0.9513753550194024|  |(10,[0,1,2,3,4,5,...| -0.4683784136986876| 0.526158102574609|  |(10,[0,1,2,3,4,5,...|-0.44652227528840105| 0.09749174131730826|  |(10,[0,1,2,3,4,5,...| 0.10157453780074743| -0.9559977576573799|  |(10,[0,1,2,3,4,5,...| 0.2105613019270259| 1.4462869745269753|  |(10,[0,1,2,3,4,5,...| 2.1214592666251364| 0.36215874190586994|  |(10,[0,1,2,3,4,5,...| 2.8497179990245116| 1.5427756079603374|  |(10,[0,1,2,3,4,5,...| 3.980473021620311| 2.949892321898203|  +--------------------+--------------------+--------------------+  only showing top 20 rows |

Run use case 200: JavaMulticlassLogisticRegressionWithElasticNetExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaMulticlassLogisticRegressionWithElasticNetExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Coefficients:  3 x 4 CSCMatrix  (1,2) -0.7666333131801358  (0,3) 0.30499989791247106  (1,3) -0.38544484160714004 |

Run Example 201: JavaMultilayerPerceptronClassifierExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaMultilayerPerceptronClassifierExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Test set accuracy = 0.9523809523809523 |

Run Example 202: JavaNaiveBayesExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaNaiveBayesExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +-----+--------------------+--------------------+-----------+----------+  |label| features| rawPrediction|probability|prediction|  +-----+--------------------+--------------------+-----------+----------+  | 0.0|(692,[95,96,97,12...|[-172664.79564650...| [1.0,0.0]| 0.0|  | 0.0|(692,[98,99,100,1...|[-176279.15054306...| [1.0,0.0]| 0.0|  | 0.0|(692,[122,123,124...|[-189600.55409526...| [1.0,0.0]| 0.0|  | 0.0|(692,[124,125,126...|[-274673.88337431...| [1.0,0.0]| 0.0|  | 0.0|(692,[124,125,126...|[-183393.03869049...| [1.0,0.0]| 0.0|  | 0.0|(692,[125,126,127...|[-256992.48807619...| [1.0,0.0]| 0.0|  | 0.0|(692,[126,127,128...|[-210411.53649773...| [1.0,0.0]| 0.0|  | 0.0|(692,[127,128,129...|[-170627.63616681...| [1.0,0.0]| 0.0|  | 0.0|(692,[127,128,129...|[-212157.96750469...| [1.0,0.0]| 0.0|  | 0.0|(692,[127,128,129...|[-183253.80108550...| [1.0,0.0]| 0.0|  | 0.0|(692,[128,129,130...|[-246528.93739632...| [1.0,0.0]| 0.0|  | 0.0|(692,[150,151,152...|[-158348.34683571...| [1.0,0.0]| 0.0|  | 0.0|(692,[152,153,154...|[-210229.50765957...| [1.0,0.0]| 0.0|  | 0.0|(692,[152,153,154...|[-242985.16248889...| [1.0,0.0]| 0.0|  | 0.0|(692,[152,153,154...|[-94622.933454005...| [1.0,0.0]| 0.0|  | 0.0|(692,[153,154,155...|[-266465.39689814...| [1.0,0.0]| 0.0|  | 0.0|(692,[153,154,155...|[-144989.71469229...| [1.0,0.0]| 0.0|  | 0.0|(692,[154,155,156...|[-283834.57437738...| [1.0,0.0]| 0.0|  | 0.0|(692,[181,182,183...|[-155256.59399829...| [1.0,0.0]| 0.0|  | 1.0|(692,[100,101,102...|[-147726.11958982...| [0.0,1.0]| 1.0|  +-----+--------------------+--------------------+-----------+----------+  only showing top 20 rows  Test set accuracy = 1.0 |

Run Example 203: JavaNGramExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaNGramExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +----------------------------------------------------------------------------------+  |ngrams |  +----------------------------------------------------------------------------------+  |[Hi I, I heard, heard about, about Spark] |  |[I wish, wish Java, Java could, could use, use case, case classes]|  |[Logistic regression, regression models, models are, are neat] |  +----------------------------------------------------------------------------------+ |

Run Example 204: JavaNormalizerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaNormalizerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +---+--------------+--------------------+  | id| features| normFeatures|  +---+--------------+--------------------+  | 0|[1.0,0.1,-8.0]|[0.10989010989010...|  | 1|[2.0,1.0,-4.0]|[0.28571428571428...|  | 2|[4.0,10.0,8.0]|[0.18181818181818...|  +---+--------------+--------------------+  +---+--------------+-------------------+  | id| features| normFeatures|  +---+--------------+-------------------+  | 0|[1.0,0.1,-8.0]|[0.125,0.0125,-1.0]|  | 1|[2.0,1.0,-4.0]| [0.5,0.25,-1.0]|  | 2|[4.0,10.0,8.0]| [0.4,1.0,0.8]|  +---+--------------+-------------------+ |

Run Example 205: JavaOneHotEncoderExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaOneHotEncoderExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +--------------+--------------+-------------+-------------+  |categoryIndex1|categoryIndex2| categoryVec1| categoryVec2|  +--------------+--------------+-------------+-------------+  | 0.0| 1.0|(2,[0],[1.0])|(2,[1],[1.0])|  | 1.0| 0.0|(2,[1],[1.0])|(2,[0],[1.0])|  | 2.0| 1.0| (2,[],[])|(2,[1],[1.0])|  | 0.0| 2.0|(2,[0],[1.0])| (2,[],[])|  | 0.0| 1.0|(2,[0],[1.0])|(2,[1],[1.0])|  | 2.0| 0.0| (2,[],[])|(2,[0],[1.0])|  +--------------+--------------+-------------+-------------+ |

Run Use Case 206: JavaOneVsRestExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaOneVsRestExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Test Error = 0.0 |

Run Example 207: JavaPCAExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaPCAExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +-------------------------------------------------------------+  |pcaFeatures|  +-------------------------------------------------------------+  |[1.6485728230883814,-4.013282700516299,-0.08551317848008963] |  |[-4.645104331781533,-1.1167972663619055,-0.08551317848009016]|  |[-6.428880535676489,-5.337951427775359,-0.08551317848009049] |  +-------------------------------------------------------------+ |

Run Use Case 208: JavaPipelineExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaPipelineExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  (4, spark ijk) --> prob=[0.6292098489668478,0.37079015103315216], prediction=0.0  (5, lmn) --> prob=[0.984770006762304,0.015229993237696027], prediction=0.0  (6, spark hadoop spark) --> prob=[0.13412348342566066,0.8658765165743394], prediction=1.0  (7, apache hadoop) --> prob=[0.9955732114398529,0.00442678856014711], prediction=0.0 |

Run Example 209: JavaPolynomialExpansionExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaPolynomialExpansionExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +----------+-----------------------------------------+  |features |polyFeatures |  +----------+-----------------------------------------+  |[2.0,1.0] |[2.0,4.0,8.0,1.0,2.0,4.0,1.0,2.0,1.0] |  |[0.0,0.0] |[0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0] |  |[3.0,-1.0]|[3.0,9.0,27.0,-1.0,-3.0,-9.0,1.0,3.0,-1.0]|  +----------+-----------------------------------------+ |

Run Use Case 210: JavaPowerIterationClusteringExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaPowerIterationClusteringExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +---+-------+  |id |cluster|  +---+-------+  |4 |1 |  |0 |0 |  |1 |0 |  |3 |1 |  |2 |0 |  +---+-------+ |

Run Example 211: JavaPrefixSpanExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaPrefixSpanExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +----------+----+  |sequence|freq|  +----------+----+  | [[1]]| 3|  | [[3]]| 2|  | [[2]]| 3|  | [[1, 2]]| 3|  |[[1], [3]]| 2|  +----------+----+ |

Run Example 212: JavaQuantileDiscretizerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaQuantileDiscretizerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +---+----+------+  |id |hour|result|  +---+----+------+  |0 |18.0|2.0 |  |1 |19.0|2.0 |  |2 |8.0 |1.0 |  |3 |5.0 |1.0 |  |4 |2.2 |0.0 |  +---+----+------+ |

Run Example 213: JavaRandomForestClassifierExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaRandomForestClassifierExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +--------------+-----+--------------------+  |predictedLabel|label| features|  +--------------+-----+--------------------+  | 0.0| 0.0|(692,[95,96,97,12...|  | 0.0| 0.0|(692,[100,101,102...|  | 0.0| 0.0|(692,[123,124,125...|  | 0.0| 0.0|(692,[123,124,125...|  | 0.0| 0.0|(692,[124,125,126...|  +--------------+-----+--------------------+  only showing top 5 rows |

Run Example 214: JavaRandomForestRegressorExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml . JavaRandomForestRegressorExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +----------+-----+--------------------+  |prediction|label| features|  +----------+-----+--------------------+  | 0.0| 0.0|(692,[121,122,123...|  | 0.05| 0.0|(692,[122,123,148...|  | 0.0| 0.0|(692,[123,124,125...|  | 0.0| 0.0|(692,[124,125,126...|  | 0.05| 0.0|(692,[124,125,126...|  +----------+-----+--------------------+  only showing top 5 rows |

Run Example 215: JavaRFormulaExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaRFormulaExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +--------------+-----+  | features|label|  +--------------+-----+  |[0.0,0.0,18.0]| 1.0|  |[1.0,0.0,12.0]| 0.0|  |[0.0,1.0,15.0]| 0.0|  +--------------+-----+ |

Run Example 216: JavaRobustScalerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaRobustScalerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +-----+--------------------+--------------------+  |label| features| scaledFeatures|  +-----+--------------------+--------------------+  | 0.0|(692,[127,128,129...|(692,[127,128,129...|  | 1.0|(692,[158,159,160...|(692,[158,159,160...|  | 1.0|(692,[124,125,126...|(692,[124,125,126...|  | 1.0|(692,[152,153,154...|(692,[152,153,154...|  | 1.0|(692,[151,152,153...|(692,[151,152,153...|  | 0.0|(692,[129,130,131...|(692,[129,130,131...|  | 1.0|(692,[158,159,160...|(692,[158,159,160...|  | 1.0|(692,[99,100,101,...|(692,[99,100,101,...|  | 0.0|(692,[154,155,156...|(692,[154,155,156...|  | 0.0|(692,[127,128,129...|(692,[127,128,129...|  | 1.0|(692,[154,155,156...|(692,[154,155,156...|  | 0.0|(692,[153,154,155...|(692,[153,154,155...|  | 0.0|(692,[151,152,153...|(692,[151,152,153...|  | 1.0|(692,[129,130,131...|(692,[129,130,131...|  | 0.0|(692,[154,155,156...|(692,[154,155,156...|  | 1.0|(692,[150,151,152...|(692,[150,151,152...|  | 0.0|(692,[124,125,126...|(692,[124,125,126...|  | 0.0|(692,[152,153,154...|(692,[152,153,154...|  | 1.0|(692,[97,98,99,12...|(692,[97,98,99,12...|  | 1.0|(692,[124,125,126...|(692,[124,125,126...|  +-----+--------------------+--------------------+  only showing top 20 rows |

Run Example 217: JavaSQLTransformerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml . JavaSQLTransformerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +---+---+---+---+----+  | id| v1| v2| v3| v4|  +---+---+---+---+----+  | 0|1.0|3.0|4.0| 3.0|  | 2|2.0|5.0|7.0|10.0|  +---+---+---+---+----+ |

Run Example 218: JavaStandardScalerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaStandardScalerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +-----+--------------------+--------------------+  |label| features| scaledFeatures|  +-----+--------------------+--------------------+  | 0.0|(692,[127,128,129...|(692,[127,128,129...|  | 1.0|(692,[158,159,160...|(692,[158,159,160...|  | 1.0|(692,[124,125,126...|(692,[124,125,126...|  | 1.0|(692,[152,153,154...|(692,[152,153,154...|  | 1.0|(692,[151,152,153...|(692,[151,152,153...|  | 0.0|(692,[129,130,131...|(692,[129,130,131...|  | 1.0|(692,[158,159,160...|(692,[158,159,160...|  | 1.0|(692,[99,100,101,...|(692,[99,100,101,...|  | 0.0|(692,[154,155,156...|(692,[154,155,156...|  | 0.0|(692,[127,128,129...|(692,[127,128,129...|  | 1.0|(692,[154,155,156...|(692,[154,155,156...|  | 0.0|(692,[153,154,155...|(692,[153,154,155...|  | 0.0|(692,[151,152,153...|(692,[151,152,153...|  | 1.0|(692,[129,130,131...|(692,[129,130,131...|  | 0.0|(692,[154,155,156...|(692,[154,155,156...|  | 1.0|(692,[150,151,152...|(692,[150,151,152...|  | 0.0|(692,[124,125,126...|(692,[124,125,126...|  | 0.0|(692,[152,153,154...|(692,[152,153,154...|  | 1.0|(692,[97,98,99,12...|(692,[97,98,99,12...|  | 1.0|(692,[124,125,126...|(692,[124,125,126...|  +-----+--------------------+--------------------+  only showing top 20 rows |

Run Example 219: JavaStopWordsRemoverExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaStopWordsRemoverExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +----------------------------+--------------------+  |raw |filtered |  +----------------------------+--------------------+  |[I, saw, the, red, balloon] |[saw, red, balloon] |  |[Mary, had, a, little, lamb]|[Mary, little, lamb]|  +----------------------------+--------------------+ |

Run Example 220: JavaStringIndexerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml . JavaStringIndexerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +---+--------+-------------+  | id|category|categoryIndex|  +---+--------+-------------+  | 0| a| 0.0|  | 1| b| 2.0|  | 2| c| 1.0|  | 3| a| 0.0|  | 4| a| 0.0|  | 5| c| 1.0|  +---+--------+-------------+ |

Run Example 221: JavaSummarizerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaSummarizerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  with weight: mean = [3.333333333333333,5.0,6.333333333333333], variance = [2.000000000000001,4.5,2.000000000000001]  without weight: mean = [3.0,4.5,6.0], variance = [2.0,4.5,2.0] |

Run Example 222: JavaTfIdfExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaTfIdfExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +-----+--------------------+  |label| features|  +-----+--------------------+  | 0.0|(20,[6,8,13,16],[...|  | 0.0|(20,[0,2,7,13,15,...|  | 1.0|(20,[3,4,6,11,19]...|  +-----+--------------------+ |

Run Example 223: JavaTokenizerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaTokenizerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +-----------------------------------+------------------------------------------+-----+  |sentence |words |tokens|  +-----------------------------------+------------------------------------------+-----+  |Hi I heard about Spark |[hi, i, heard, about, spark] |5 |  |I wish Java could use case classes |[i, wish, java, could, use, case, classes]|7 |  |Logistic,regression,models,are,neat|[logistic,regression,models,are,neat] |1 |  +-----------------------------------+------------------------------------------+-----+  +-----------------------------------+------------------------------------------+-----+  |sentence |words |tokens|  +-----------------------------------+------------------------------------------+-----+  |Hi I heard about Spark |[hi, i, heard, about, spark] |5 |  |I wish Java could use case classes |[i, wish, java, could, use, case, classes]|7 |  |Logistic, regression, models, are, neat|[logistic, regression, models, are, neat] |5 |  +-----------------------------------+------------------------------------------+-----+ |

Run Example 224: JavaUnivariateFeatureSelectorExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaUnivariateFeatureSelectorExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +---+--------------------+-----+----------------+  | id| features|label|selectedFeatures|  +---+--------------------+-----+----------------+  | 1|[1.7,4.4,7.6,5.8,...| 3.0| [2.3]|  | 2|[8.8,7.3,5.7,7.3,...| 2.0| [4.1]|  | 3|[1.2,9.5,2.5,3.1,...| 3.0| [2.5]|  | 4|[3.7,9.2,6.1,4.1,...| 2.0| [3.8]|  | 5|[8.9,5.2,7.8,8.3,...| 4.0| [3.0]|  | 6|[7.9,8.5,9.2,4.0,...| 4.0| [2.1]|  +---+--------------------+-----+----------------+ |

Run Example 225: JavaVarianceThresholdSelectorExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaVarianceThresholdSelectorExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +---+--------------------+-----------------+  | id| features| selectedFeatures|  +---+--------------------+-----------------+  | 1|[6.0,7.0,0.0,7.0,...|[6.0,0.0,7.0,0.0]|  | 2|[0.0,9.0,6.0,0.0,...|[0.0,6.0,0.0,9.0]|  | 3|[0.0,9.0,3.0,0.0,...|[0.0,3.0,0.0,5.0]|  | 4|[0.0,9.0,8.0,5.0,...|[0.0,8.0,5.0,4.0]|  | 5|[8.0,9.0,6.0,5.0,...|[8.0,6.0,5.0,4.0]|  | 6|[8.0,9.0,6.0,0.0,...|[8.0,6.0,0.0,0.0]|  +---+--------------------+-----------------+ |

Run Example 226: JavaVectorAssemblerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaVectorAssemblerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +-----------------------+-------+  |features |clicked|  +-----------------------+-------+  |[18.0,1.0,0.0,10.0,0.5]|1.0 |  +-----------------------+-------+ |

Run Example 227: JavaVectorIndexerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaVectorIndexerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +-----+--------------------+--------------------+  |label| features| indexed|  +-----+--------------------+--------------------+  | 0.0|(692,[127,128,129...|(692,[127,128,129...|  | 1.0|(692,[158,159,160...|(692,[158,159,160...|  | 1.0|(692,[124,125,126...|(692,[124,125,126...|  | 1.0|(692,[152,153,154...|(692,[152,153,154...|  | 1.0|(692,[151,152,153...|(692,[151,152,153...|  | 0.0|(692,[129,130,131...|(692,[129,130,131...|  | 1.0|(692,[158,159,160...|(692,[158,159,160...|  | 1.0|(692,[99,100,101,...|(692,[99,100,101,...|  | 0.0|(692,[154,155,156...|(692,[154,155,156...|  | 0.0|(692,[127,128,129...|(692,[127,128,129...|  | 1.0|(692,[154,155,156...|(692,[154,155,156...|  | 0.0|(692,[153,154,155...|(692,[153,154,155...|  | 0.0|(692,[151,152,153...|(692,[151,152,153...|  | 1.0|(692,[129,130,131...|(692,[129,130,131...|  | 0.0|(692,[154,155,156...|(692,[154,155,156...|  | 1.0|(692,[150,151,152...|(692,[150,151,152...|  | 0.0|(692,[124,125,126...|(692,[124,125,126...|  | 0.0|(692,[152,153,154...|(692,[152,153,154...|  | 1.0|(692,[97,98,99,12...|(692,[97,98,99,12...|  | 1.0|(692,[124,125,126...|(692,[124,125,126...|  +-----+--------------------+--------------------+  only showing top 20 rows |

Run Example 228: JavaVectorSizeHintExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaVectorSizeHintExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Rows where 'userFeatures' is not the right size are filtered out  +---+----+------+--------------+-------+  |id |hour|mobile|userFeatures |clicked|  +---+----+------+--------------+-------+  |0 |18 |1.0 |[0.0,10.0,0.5]|1.0 |  +---+----+------+--------------+-------+  Assembled columns 'hour', 'mobile', 'userFeatures' to vector column 'features'  +-----------------------+-------+  |features |clicked|  +-----------------------+-------+  |[18.0,1.0,0.0,10.0,0.5]|1.0 |  +-----------------------+-------+ |

Run Example 229: JavaVectorSlicerExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaVectorSlicerExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  +--------------------+-------------+  |userFeatures |features |  +--------------------+-------------+  |(3,[0,1],[-2.0,2.3])|(2,[0],[2.3])|  |[-2.0,2.3,0.0] |[2.3,0.0] |  +--------------------+-------------+ |

Run Example 230: JavaWord2VecExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml.JavaWord2VecExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.ml .xxx  # ARM output  Text: [Hi, I, heard, about, Spark] =>  Vector: [-0.02431743445340544,-0.015829450637102126,0.01689382940530777]  Text: [I, wish, Java, could, use, case, classes] =>  Vector: [-0.05370100228382008,-0.018913335033825463,0.023755997818495543]  Text: [Logistic, regression, models, are, neat] =>  Vector: [0.01655797511339188,-0.027168981288559737,0.08163769766688347] |

1. Java test cases: test cases in the mllib/ folder

Run Case 231: JavaALS

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaALS --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaALS.  # ARM output  Usage: JavaALS <ratings\_file> <rank> <iterations> <output\_dir> [<blocks>] |

Run Example 232: JavaAssociationRulesExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaAssociationRulesExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaAssociationRulesExample.  # ARM output  [ a] => [ b ], 0.8 |

Run Example 233: JavaBinaryClassificationMetricsExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaBinaryClassificationMetricsExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaBinaryClassificationMetricsExample  # ARM output  Precision by threshold: [(0.9999999999999272,1.0), (0.99999999999992477,1.0), (0.99999999999984444,1.0), (0.99999999999959581,1.0), (0.9999999999945857,1.0), (0.9999999999890732,1.0), (0.9999999999623332,1.0), (0.9999999996994677,1.0), (0.9999999996333453,1.0), (0.9999999994604953,1.0), (0.9999999994336923,1.0), (0.99999999991649884,1.0), (0.99999999990747106,1.0), (0.99999999989195318,1.0), (0.99999999987163963,1.0), (0.99999999981564602,1.0), (0.99999999972193769,1.0), (0.99999999968315649,1.0), (0.99999999544786299,1.0), (0.9999993899615596,1.0), (0.9999964985848224,1.0), (0.12211587409472682,0.9545454545454546), (0.048984149073111535,0.9565217391304348), (0.0012853401732941941,0.9166666666666666), (3.818068237499681E-5,0.88), (4.835863200872565E-7,0.8461538461538461), (4.793413812601818E-7,0.8148148148148148), (2.520749065679303E-7,0.7857142857142857), (1.1777070310078065E-7,0.7586206896551724), (1.9911411445086488E-9,0.7333333333333333), (1.326952551335953E-9,0.7096774193548387), (6.055547165943601E-10,0.6875), (4.83383282606745E-10,0.6666666666666666), (4.800906233086011E-10,0.6470588235294118), (9.78497592468649E-11,0.6285714285714286), (2.945905946899788E-11,0.6111111111111112), (2.4997812941187832E-11,0.5945945945945946), (7.818085626819181E-12,0.5789473684210527), (1.513848295295787E-13,0.5641025641025641), (4.696925106679965E-14,0.55)] |

Run example 234: mllib\_JavaBisectingKMeansExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaBisectingKMeansExample. --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib . JavaBisectingKMeansExample  # ARM output  Compute Cost: 0.15999999999999986  Cluster Center 0: [0.2,0.2]  Cluster Center 1: [10.2,10.2]  Cluster Center 2: [20.200000000000003,20.200000000000003]  Cluster Center 3: [30.200000000000003,30.200000000000003] |

Run example 235: mllib\_JavaChiSqSelectorExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaChiSqSelectorExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaChiSqSelectorExample  # ARM output  filtered data:  Exception in thread "main" org.apache.spark.SparkException: Task not serializable |

Run Example 236: mllib\_JavaCorrelationsExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaCorrelationsExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaCorrelationsExample  # ARM output  Correlation is: 0.8500286768773004  1.0 0.9788834658894731 0.9903895695275673  0.9788834658894731 1.0 0.9977483233986101  0.9903895695275673 0.9977483233986101 1.0 |

Run example 237: mllib\_JavaDecisionTreeClassificationExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaDecisionTreeClassificationExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaDecisionTreeClassificationExample  # ARM output  Exception in thread "main" java.lang.IllegalAccessException: Class org.apache.spark.deploy.JavaMainApplication can not access a member of class org.apache.spark.examples.mllib.JavaDecisionTreeClassificationExample with modifiers "public static" |

Run Example 238: mllib\_JavaDecisionTreeRegressionExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaDecisionTreeRegressionExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaDecisionTreeRegressionExample  # ARM output  Exception in thread "main" java.lang.IllegalAccessException: Class org.apache.spark.deploy.JavaMainApplication can not access a member of class org.apache.spark.examples.mllib.JavaDecisionTreeRegressionExample with modifiers "public static" |

Run example 239: mllib\_JavaElementwiseProductExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaElementwiseProductExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaElementwiseProductExample  # ARM output  transformedData:  Exception in thread "main" org.apache.spark.SparkException: Task not serializable |

Run example 240: mllib\_JavaGaussianMixtureExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaGaussianMixtureExample  --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaGaussianMixtureExample  # ARM output  Exception in thread "main" org.apache.hadoop.mapred.InvalidInputException: Input path does not exist: file:/home/zhuchenhao/spark-3.5.1/target/org.apache.spark.JavaGaussianMixtureExample/GaussianMixtureModel/metadata |

Run Example 241: JavaGradientBoostingClassificationExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaGradientBoostingClassificationExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaGradientBoostingClassificationExample  # ARM output  Test Error: 0.0  Learned classification GBT model:  TreeEnsembleModel classifier with 3 trees  Tree 0:  If (feature 434 <= 70.5)  If (feature 100 <= 193.5)  Predict: -1.0  Else (feature 100 > 193.5)  Predict: 1.0  Else (feature 434 > 70.5)  Predict: 1.0  Tree 1:  If (feature 434 <= 70.5)  If (feature 344 <= 253.5)  Predict: -0.47681168808847024  Else (feature 344 > 253.5)  Predict: 0.4768116880884694  Else (feature 434 > 70.5)  Predict: 0.47681168808847024  Tree 2:  If (feature 434 <= 70.5)  If (feature 344 <= 253.5)  Predict: -0.4381935810427206  Else (feature 344 > 253.5)  Predict: 0.43819358104271977  Else (feature 434 > 70.5)  Predict: 0.4381935810427206 |

Run Example 242: JavaGradientBoostingRegressionExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaGradientBoostingRegressionExample--master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaAssociationRulesExample.  # ARM output  Test Mean Squared Error: 0.037037037037037035  Learned regression GBT model:  TreeEnsembleModel regressor with 3 trees  Tree 0:  If (feature 406 <= 126.5)  If (feature 100 <= 193.5)  Predict: 0.0  Else (feature 100 > 193.5)  Predict: 1.0  Else (feature 406 > 126.5)  Predict: 1.0  Tree 1:  Predict: 0.0  Tree 2:  Predict: 0.0 |

Run Example 243: JavaHypothesisTestingExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaHypothesisTestingExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaAssociationRulesExample.  # ARM output  Chi squared test summary:  method: pearson  degrees of freedom = 4  statistic = 0.12499999999999999  pValue = 0.998126379239318  No presumption against null hypothesis: observed follows the same distribution as expected..  Chi squared test summary:  method: pearson  degrees of freedom = 2  statistic = 0.14141414141414144  pValue = 0.931734784568187  No presumption against null hypothesis: the occurrence of the outcomes is statistically independent..  Column 1:  Chi squared test summary:  method: pearson  degrees of freedom = 1  statistic = 3.0000000000000004  pValue = 0.08326451666354884  Low presumption against null hypothesis: the occurrence of the outcomes is statistically independent..  Column 2:  Chi squared test summary:  method: pearson  degrees of freedom = 1  statistic = 0.75  pValue = 0.3864762307712326  No presumption against null hypothesis: the occurrence of the outcomes is statistically independent..  Column 3:  Chi squared test summary:  method: pearson  degrees of freedom = 2  statistic = 3.0  pValue = 0.22313016014843035  No presumption against null hypothesis: the occurrence of the outcomes is statistically independent.. |

Run Example 244: JavaHypothesisTestingKolmogorovSmirnovTestExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaHypothesisTestingKolmogorovSmirnovTestExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaAssociationRulesExample.  # ARM output  Kolmogorov-Smirnov test summary:  degrees of freedom = 0  statistic = 0.539827837277029  pValue = 0.06821463111921133  Low presumption against null hypothesis: Sample follows theoretical distribution. |

Run example 245: mllib\_JavaIsotonicRegressionExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaIsotonicRegressionExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaIsotonicRegressionExample  # ARM output  Mean Squared Error = 0.007292013545930122 |

Run Example 246: JavaKernelDensityEstimationExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaKernelDensityEstimationExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaAssociationRulesExample.  # ARM output  [0.04145944023341911, 0.07902016933085627, 0.08962920127312336] |

Run Example 247: mllin\_JavaKMeansExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaKMeansExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaKMeansExample  # ARM output  Cluster centers:  [9.1,9.1,9.1]  [0.1,0.1,0.1]  Cost: 0.11999999999999996  Within Set Sum of Squared Errors = 0.1199999999999996 |

Run Example 248: JavaLatentDirichletAllocationExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaLatentDirichletAllocationExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaAssociationRulesExample.  # ARM output  Learned topics (as distributions over vocab of 11 words):  Topic 0: 8.609045517179787 9.363310760965403 6.122700081611796 23.673017104912965 4.794869231879549 3.9111499074509823 8.94613036455791 1.4710483628540962 2.135164174751737 6.230094517723791 9.979571220509177  Topic 1: 4.434139201807829 10.719266109890903 3.2411103616312205 8.50104807565507 3.006363865146792 3.073573897935561 14.351432955048711 0.7367903591538527 4.5217575624867274 13.441197057862873 21.664564303099773  Topic 2: 12.956815281012386 8.917423129143694 2.636189556756984 7.825934819431966 17.19876690297366 15.015276194613456 7.702436680393381 7.792161277992051 1.3430782627615356 4.328708424413336 1.3558644763910517 |

Run Example 249: JavaLBFGSExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaLBFGSExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaLBFGSExample  # ARM output  24/08/03 03:03:43 ERROR LBFGS: Failure! Resetting history: breeze.optimize.FirstOrderException: Line search failed  Loss of each step in training process  0.6931471805599458  0.33430243421184935  0.19473095533260118  0.0020928873302658224  7.100781251966798E-4  7.100781251966798E-4  5.751637018492142E-4  9.267685208079032E-5  5.772617267354478E-5  5.550371488281082E-5  4.4114776678047625E-5  3.4942151536055275E-5  3.142697044086399E-5  2.66651045013966E-5  2.422321597152757E-5  2.3573285798413956E-5  2.327347665711998E-5  2.2982783580032414E-5  2.2830177765411474E-5  2.265404121974377E-5  2.2515791851525292E-5  2.2500163487917206E-5  Area under ROC = 1.0 |

Run Example 250: JavaLogisticRegressionWithLBFGSExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib.JavaLogisticRegressionWithLBFGSExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaLogisticRegressionWithLBFGSExample  # ARM output  Accuracy = 1.0 |

Run Example 251: JavaMulticlassClassificationMetricsExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaMulticlassClassificationMetricsExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaMulticlassClassificationMetricsExample  # ARM output  Confusion matrix:  15.0 0.0 2.0  0.0 19.0 0.0  7.0 0.0 16.0  Accuracy = 0.847457627118644  Class 0.000000 precision = 0.681818  Class 0.000000 recall = 0.882353  Class 0.000000 F1 score = 0.769231  Class 1.000000 precision = 1.000000  Class 1.000000 recall = 1.000000  Class 1.000000 F1 score = 1.000000  Class 2.000000 precision = 0.888889  Class 2.000000 recall = 0.695652  Class 2.000000 F1 score = 0.780488  Weighted precision = 0.865006  Weighted recall = 0.847458  Weighted F1 score = 0.847935  Weighted false positive rate = 0.069680 |

Run Example 252: JavaMultiLabelClassificationMetricsExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaMultiLabelClassificationMetricsExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaMultiLabelClassificationMetricsExample  # ARM output  Recall = 0.642857  Precision = 0.666667  F1 measure = 0.638095  Accuracy = 0.547619  Class 0.0 precision = 1.000000  Class 0.0 recall = 0.800000  Class 0.0 F1 score = 0.888889  Class 1.0 precision = 0.666667  Class 1.0 recall = 0.666667  Class 1.0 F1 score = 0.666667  Micro recall = 0.666667  Micro precision = 0.727273  Micro F1 measure = 0.695652  Hamming loss = 0.333333  Subset accuracy = 0.285714 |

Run example 253: mllib\_JavaNaiveBayesExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaNaiveBayesExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaNaiveBayesExample  # ARM output  Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory file:/home/zhuchenhao/spark-3.5.1/target/tmp/myNaiveBayesModel/metadata already exists |

According to the ARM error, there is already an output model in the output path, which needs to be deleted before running the use case:   
Delete file:

|  |
| --- |
| rm -rf spark-3.5.1/target/tmp/myNaiveBayesModel |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaNaiveBayesExample  # ARM output  <Successful run with no command line output> |

Run example 254: mllib\_JavaPCAExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaPCAExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaPCAExample  # ARM output  Projected vector of principal component:  [1.6485728230883814,-4.013282700516299,-0.08551317848008963,2.487286050101037]  [-4.645104331781532,-1.1167972663619055,-0.08551317848008999,2.487286050101038]  [-6.42888053567649,-5.337951427775359,-0.08551317848009043,2.487286050101037] |

Run example 255: mllib\_JavaPowerIterationClusteringExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaPowerIterationClusteringExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaPowerIterationClusteringExample  # ARM output  4 -> 1  0 -> 0  1 -> 0  3 -> 0  5 -> 1  2 -> 0 |

Run example 256: mllib\_JavaPrefixSpanExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaPrefixSpanExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaPrefixSpanExample  # ARM output  [[twenty three  [[3]], 2  [[1]], 3  [[2, 1]], 3  [[1], [3]], 2 |

Run example 257: mllib\_JavaRandomForestClassificationExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaRandomForestClassificationExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaRandomForestClassificationExample  # ARM output  Test Error: 0.033333333333333333  Learned classification forest model:  TreeEnsembleModel classifier with 3 trees  Tree 0:  If (feature 540 <= 10.0)  If (feature 274 <= 6.0)  If (feature 551 <= 4.5)  Predict: 1.0  Else (feature 551 > 4.5)  Predict: 0.0  Else (feature 274 > 6.0)  Predict: 0.0  Else (feature 540 > 10.0)  Predict: 0.0  Tree 1:  If (feature 406 <= 126.5)  Predict: 0.0  Else (feature 406 > 126.5)  Predict: 1.0  Tree 2:  If (feature 433 <= 52.5)  Predict: 0.0  Else (feature 433 > 52.5)  Predict: 1.04.328708424413336 1.3558644763910517 |

Run Example 258: mllib\_JavaRandomForestRegressionExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaRandomForestRegressionExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaRandomForestRegressionExample  # ARM output  Learned topics (as distributions over vocab of 11 words):  Test Mean Squared Error: 0.0  Learned regression forest model:  TreeEnsembleModel regressor with 3 trees  Tree 0:  If (feature 490 <= 43.0)  Predict: 0.0  Else (feature 490 > 43.0)  Predict: 1.0  Tree 1:  If (feature 406 <= 9.5)  Predict: 0.0  Else (feature 406 > 9.5)  If (feature 538 <= 2.5)  Predict: 1.0  Else (feature 538 > 2.5)  Predict: 0.0  Tree 2:  If (feature 462 <= 63.0)  If (feature 407 <= 83.0)  Predict: 0.0  Else (feature 407 > 83.0)  Predict: 1.0  Else (feature 462 > 63.0)  Predict: 1.0 |

Run Example 259: JavaRankingMetricsExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaRankingMetricsExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaRankingMetricsExample  # ARM output  Precision at 1 = 0.300000  NDCG at 1 = 0.300000  Recall at 1 = 0.030000  Precision at 3 = 0.422222  NDCG at 3 = 0.394227  Recall at 3 = 0.126667  Precision at 5 = 0.446667  NDCG at 5 = 0.418686  Recall at 5 = 0.223333  Mean average precision = 0.257974  Mean average precision at 2 = 0.300000  RMSE = 0.230519  R-squared = 0.959295 |

Run Example 260: JavaRecommendationExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaRecommendationExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaRecommendationExample  # ARM output  Mean Squared Error = 7.263540590218692E-6 |

Run Use Case 261: JavaSimpleFPGrowth

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaSimpleFPGrowth --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaSimpleFPGrowth  # ARM output  [[z]], 5  [[x]], 4  [[x, z]], 3  [[y]], 3  [[y, x]], 3  [[y, x, z]], 3  [[y, z]], 3  [[r]], 3  [[r, x]], 2  [[r, z]], 2  [[s]], 3  [[s, y]], 2  [[s, y, x]], 2  [[s, y, x, z]], 2  [[s, y, z]], 2  [[s, x]], 3  [[s, x, z]], 2 |

Run Example 262: JavaStratifiedSamplingExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaStratifiedSamplingExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaStratifiedSamplingExample  # ARM output  approxSample size is 2  1 a  2d  exactSample size is 4  1 a  2d  2 e  3f |

Run Example 263: JavaStreamingTestExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaStreamingTestExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaAssociationRulesExample.  # ARM output  Usage: JavaStreamingTestExample <dataDir> <batchDuration> <numBatchesTimeout> |

Run Example 264: JavaSummaryStatisticsExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaSummaryStatisticsExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaSummaryStatisticsExample  # ARM output  [2.0,20.0,200.0]  [1.0,100.0,10000.0]  [3.0,3.0,3.0] |

Run Example 265: JavaSVDExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaSVDExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaSVDExample  # ARM output  U factor is:  [-0.3882913051166563,-0.9198099362554474,-0.05638744130170942,-2.3748725652694702E-8,1.4901161193847656E-8]  [-0.5301719995198348,0.27301855119012264,-0.8027319114319467,-7.450580596923828E-9,2.3283064365386963E-8]  [-0.7537556058139431,0.28179877904596407,0.5936682026454336,-5.9604644775390625E-8,-1.5832483768463135E-8]  Singular values are: [13.029275535600474,5.368578733451684,2.533049821881375,4.6944519248162234E-8,1.615330408456659E-8]  V factor is:  -0.3127853433723262 0.31167135691578307 0.3036691110743833 ... (5 total)  -0.02980145013095395 -0.1713321126360875 -0.022260691761612522 ...  -0.12207248163673154 0.15256470925290191 -0.9507099755768705 ...  -0.7184789931874107 -0.6809628499946365 -0.017224500912004143 ...  -0.6084105917199363 0.6217072292290714 0.056065956591734765 ... |

Run Example 266: JavaSVMWithSGDExample

|  |
| --- |
| ./bin/spark-submit --class org.apache.spark.examples.ml lib .JavaSVMWithSGDExample --master local[1] examples/target/original-spark-examples\_2.12-3.5.1.jar |

Operation results:

|  |
| --- |
| # RISC-V output  Error: Failed to load class org.apache.spark.examples.mllib.JavaSVMWithSGDExample  # ARM output  Area under ROC = 1.0 |

1. **Perf performance analysis**
2. Python automated testing script uses:

|  |
| --- |
| # Note that you must run the script in the directory above spark-3.5.1 (assuming the directory above is ~), and the perf\_bench .sh script needs to be placed in the spark-3.5.1 folder directory  cd ~  nohup bash run\_ml\_python.sh  # The running results will be saved in the nohup.out file in the ~ directory |

There are 6 Python automation scripts, namely run\_ml\_python.sh , run\_mllib\_python.sh , run\_pyspark\_ml\_connect\_python.sh , run\_pyspark\_ml\_python.sh , run\_pyspark\_mllib\_python.sh , run\_pyspark\_ml\_tuning\_python.sh , all of which must be run according to the above usage.

2. Java automated test script usage:

|  |
| --- |
| # Note that you must run the script in the directory above spark-3.5.1 (assuming the directory above is ~), and the perf\_bench .sh script needs to be placed in the spark-3.5.1 folder directory  cd ~  nohup bash run\_ml\_java.sh  # The running results will be saved in the nohup.out file in the ~ directory |

There are two Java automation scripts, run\_ml\_java.sh and run\_mllib\_java.sh