pm3140_proj

August 28, 2021

```python
[1]: from pyspark.sql import SparkSession

     # Spark session & context
     spark = SparkSession.builder.master('local[2]').getOrCreate()
     sc = spark.sparkContext
```

```
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform
(file:/usr/local/spark-3.1.2-bin-hadoop3.2/jars/spark-unsafe_2.12-3.1.2.jar) to
constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of
org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal
reflective access operations
WARNING: All illegal access operations will be denied in a future release
21/08/28 00:29:22 WARN NativeCodeLoader: Unable to load native-hadoop library
for your platform… using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use
setLogLevel(newLevel).
```

```python
[2]: spark
```

```
[2]: <pyspark.sql.session.SparkSession at 0x7f5b0ae2e070>
```

```python
[3]: sc
```

```
[3]: <SparkContext master=local[2] appName=pyspark-shell>
```

```python
[5]: spark.conf.set("spark.sql.caseSensitive", "true")
```

0.1 Reading data

```python
[6]: df_artists1 = spark.read.json("dataset-full-ratings/artists/all_artists.json")
     df_artists = df_artists1.select('_id','name')
     df_artists.show(truncate=False)
```

```
df_artists.count()
## artists list
```

```
+----------------------------------+--------------------+
|_id                               |name                |
+----------------------------------+--------------------+
|00010eb3-ebfe-4965-81ef-0ac64cd49fde|La Niña de los Peines|
|00034ede-a1f1-4219-be39-02f36853373e|O Rappa             |
|0003fd17-b083-41fe-83a9-d550bd4f00a1|                    |
|0004537a-4b12-43eb-a023-04009e738d2e|Ultra Naté          |
|0013bcdd-fe35-4c9f-ac41-4b41b9000e17|Siobhan Magnus      |
|00302a51-04f5-4b2d-8993-fcd1c2ae1cf4|                    |
|0035a150-bceb-4abe-88ca-a494cdf14968|Laura McCormick     |
|0035c656-9853-44a8-b105-c44089a43cea|Diandra             |
|00376321-ce0f-4bd7-a98f-fcabdbf06ea7|Raappana            |
|0039c7ae-e1a7-4a7d-9b49-0cbc716821a6|Death Cab for Cutie |
|003b2747-b74a-46c1-a51e-aeaffe88256c|Erdmöbel            |
|0040c89b-f2e6-4bc3-b75d-5152fb0c890e|Bernhard Romberg    |
|00413ec4-1dd2-4878-b98a-743cf6499501|Rauno Lehtinen      |
|0042aa4d-a972-4a2a-b7cf-8044d09bbc67|Andrew Fletcher     |
|00445d42-1e63-48e7-ba2f-066cd37a0927|Hilding Rosenberg   |
|004e5eed-e267-46ea-b504-54526f1f377d|The Gathering       |
|004e6286-a1c3-4174-b9f5-710666e39ecf|Richie Loop         |
|005a1712-6d74-47cb-9032-8bd63febd966|Nichole Cordova     |
|005dd8c0-1c1c-4ed2-8963-4249449f5901|Ernst Jandl         |
|006631f4-5214-49ff-a386-a064853d1a1e|Marco Antonio Solís |
+----------------------------------+--------------------+
only showing top 20 rows
```

[6]: 10281

```
[7]: import json
     import re

     def return_df_json(inp_path):

         with open(inp_path) as file:
             json_string = file.read()

         json_objects = re.sub('}{', '}|; |{', json_string).split('|; |')

         obj_list = []
         for json_object in json_objects:
             obj_list.append(json.loads(json_object))
```

```
    df = spark.createDataFrame(obj_list)
    return df
```

[8]:
```
df_artists_ratings = return_df_json("dataset-full-ratings/artists_ratings/
 ↪all_ratings_artists.json")
df_artists_ratings.show(n=10,truncate = False)
## the artist rating given by the users
```

21/08/28 00:29:38 WARN TaskSetManager: Stage 4 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
[Stage 4:>   (0 + 1) / 1]

```
+------------------------------------+------------------------------------+-----
-+----------+
|_id                                 |document_id                         
|rating|user_id   |
+------------------------------------+------------------------------------+-----
-+----------+
|886edc32-1289-4de8-9b84-f3f68a4c2244|655f9089-685f-44cf-baec-fda740adabde|1.5
|qkuphal   |
|46142ad5-e62e-4d30-8e41-defe6040ab39|c8c021ad-6397-4990-b3b7-23d027c5931a|1.0
|gkeebler  |
|7153523a-0cc0-4182-9876-aa95c169aeb9|1ac7bca8-b96a-4888-9de5-6f0a71182230|3.0
|akovacek  |
|ee08a2a9-57b0-4281-abd6-78515502e1cc|655f9089-685f-44cf-baec-fda740adabde|4.5
|gkoepp    |
|05038a35-1f56-4baa-a106-c5c97c906ad2|0b55e988-e2e5-458d-9269-bb8c54825466|3.5
|proob     |
|cb65a323-7cdd-43fd-af33-e5194848d88b|b63be4a2-51e3-472a-955f-f2c74870ba1d|4.0
|lhayes    |
|73ca99de-0d4a-43d3-a580-986dfbe8cab8|7113caec-65cd-439c-8dbc-089b8684d3f9|4.0
|araynor   |
|98b93102-2149-4cd8-93e2-584a032544d2|4d2df7be-3eb1-4a96-abd7-5db8918d4494|4.0
|mturner   |
|fb723a4d-b29f-4af6-a0ac-abeb937088d4|f115787a-99d2-48a3-b249-fb90aba48d3a|1.5
|arice     |
|6eaf216e-9f3d-4d0b-a1be-d07cc9b9dbee|3f27316a-ab8b-488f-a1c8-0584473a27e5|1.0
|hgreenholt|
+------------------------------------+------------------------------------+-----
-+----------+
only showing top 10 rows
```

[9]:
```
df_albums1 = return_df_json("dataset-full-ratings/albums/all_albums.json")
```

[10]:
```
df_albums = df_albums1.select('_id','name');
```

```
[11]: df_albums.show(truncate=False) ## album list
```

```
+------------------------------------+-----------------------------+
|_id                                 |name                         |
+------------------------------------+-----------------------------+
|00031241-434d-4f54-b170-f64db965e1fe|No I in Threesome            |
|0010ff0b-eb27-42e5-b1ad-c237f1e44b63|Breathe                      |
|0014a89f-978c-401f-b3cb-86d14d41ea0d|Digital Vein                 |
|00175090-3a98-4b0c-8714-0160af81588d|Ara Sicak                    |
|0018507e-15b4-4835-b319-e5ff0a89dbd7|Stupid Summer Dreams         |
|001a02f5-67aa-4ee9-a408-949ae60fc5cf|Bulimia Sisters              |
|0021f266-489a-33a8-b91f-506e37b6385d|Split                        |
|002abecb-2878-46e2-a9d8-e4b75edd4b0b|                             |
|002de762-fffa-43cd-9346-46ec581cc7b6|Death At My Door             |
|00327e39-61f8-48b6-ab7d-0d374182ce1b|Niin tuleni teen             |
|003422d5-70f2-4c12-af55-9f4a9f804498|                             |
|003f2fbd-8b2a-4244-aae7-0ee0a1d03438|Missing Pieces               |
|005e2e79-00f3-3a98-9a7d-9a1ca42de150|Leitmotif                    |
|006eeb56-6f1d-4950-af5f-31d5428bd594|Tears Don't Fall             |
|0072e313-0248-3871-8170-e2d996cea861|Reservoir                    |
|007ddfdc-f60b-4900-93c1-03410ddc5813|Mammoth                      |
|008e0463-cf49-4af8-930d-096d3e502510|Patriotic Shock              |
|00949f48-edc5-4c9d-a77e-141f5fff84e8|The Carnival (Notting Hill)  |
|009a9b2f-9d12-4adc-a4ff-b373a5aae0c9|The Enchanted Wood           |
|00a6c638-2200-4543-bac5-65b8d33469fa|                             |
+------------------------------------+-----------------------------+
only showing top 20 rows
```

21/08/28 00:29:41 WARN TaskSetManager: Stage 5 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.

```
[12]: df_albums_ratings = return_df_json("dataset-full-ratings/albums_ratings/
      ↪all_ratings_albums.json")
```

```
[13]: df_albums_ratings.orderBy('_id').show(truncate= False) ## album ratings
```

21/08/28 00:29:45 WARN TaskSetManager: Stage 6 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.

```
+------------------------------------+-----------------------------------+-----
-+---------+
|_id                                 |document_id                        
|rating|user_id  |
+------------------------------------+-----------------------------------+-----
-+---------+
|0000d95f-e2db-400b-b1a4-53f4461f3c68|90f0350d-d10a-429c-83dc-a6b674771f01|1.0
|aschimmel|
|00016ed8-515a-40d6-b0ee-5ccd1b43cc4f|6f2178e9-31d1-33db-acbb-2a31c85e65d6|3.5
```

```
|aschmitt |
|0002c72c-0b91-407d-b8d1-2c889914f866|83f64ffa-c945-316b-bd16-10ad714ec334|1.0
|mhahn     |
|0003afe0-b6d4-49d5-bb85-384270503953|1c3bbc4d-be4d-45d3-b49a-d612c0385cd2|5.0
|gjohnson  |
|0003d2ff-97af-49ef-bbe9-8abadd941b91|ea7ef450-93d8-46a3-a884-5ccb4fcc4f3a|4.5
|phyatt    |
|0003e3b0-1aea-42d9-9d45-7ca61111faa4|600cc060-b965-4909-a017-2da06972ab4d|1.5
|akovacek  |
|0003e8b8-c766-4c34-817c-c5f1e4be1859|ef30992a-1653-39e9-96f7-cc802e8d6a55|4.0
|rzieme    |
|000490a1-7e89-4091-b958-f2ab998f2b78|d5edfb8b-b206-48e8-9bc3-f85f41575b1b|3.5
|flabadie  |
|0004d6ee-da7a-4c87-bd9c-44ec9068ac69|e7e60bcd-f610-4cd4-8dd4-700e4a27fb2f|3.5
|phyatt    |
|00051472-ec0d-443c-af18-beaec44216b9|e23a936c-eda9-45ba-8a3f-9b68bc0b993c|4.5
|fgaylord  |
|00059e14-7687-4684-9097-1d4b7d644512|b22e4727-d9f8-468e-a463-ff25695fdf08|2.5
|aschmitt  |
|0005a168-2b66-416b-979f-6a731d4148dd|39858f32-d9f0-41e1-ad92-aabc0740bcb1|3.0
|plynch    |
|0005b1ab-57b0-472f-a427-b129437f589a|735b5038-21b8-4722-ac12-7fe08b1b72c2|4.0
|rkuhn     |
|000638c0-a243-43fb-be06-30477b4d91f0|02a2a0d4-d98b-41f1-be85-139d116b6ed1|2.5
|ccarter   |
|00069f1d-b7eb-481a-972b-67870c0b6c14|f4ef7476-386a-4d39-a76f-dc41e92fffd5|0.5
|gstamm    |
|0006b5e1-28f5-4b7d-ba65-8ee9ed6658bc|08844131-a85f-461c-925d-db6c54a25c96|4.5
|eflatley  |
|0006c0d6-cf1b-42d3-a63d-9811a09b61e8|1f49ee82-34dd-4ce1-9929-30584ed59bab|1.5
|lvolkman  |
|000731ea-4311-4a92-859c-bfadcb60aa90|cdafeea8-a1db-4ed9-8a5d-1f1be9c65583|5.0
|mkshlerin|
|0007ba63-cdd5-4fa9-a124-db6bf0f4d104|0bb622dc-c719-4c1b-8d9b-33063557b998|5.0
|bdare     |
|00082ff7-3143-4b6d-9300-9a73e7e0afb2|1a3dab35-b315-4916-b09c-7da62d39d6ef|5.0
|ksmith    |
+--------------------------------+--------------------------------+-----
--+---------+
only showing top 20 rows
```

[14]: `df_users = spark.read.json('dataset-full-ratings/users/*')` ## *user list*

[15]: `df_users.show()`

```
+-------------+----------+-------------+
|          _id|first_name|    last_name|
+-------------+----------+-------------+
|srunolfsdottir|   Seamus|Runolfsdottir|
|  cbartoletti| Celestine|   Bartoletti|
|  kschamberger|   Kattie|  Schamberger|
|  sschamberger|   Sidney|  Schamberger|
|  vschamberger|   Victor|  Schamberger|
|   dmacejkovic|  Delfina|   Macejkovic|
| rchristiansen|      Roy| Christiansen|
|    eankunding| Everette|     Ankunding|
|    erutherford|   Emilio|   Rutherford|
|   evandervort|   Elvera|   Vandervort|
|    gankunding| Gregoria|     Ankunding|
|   hbartoletti|   Hector|   Bartoletti|
|  hcartwright|   Hailee|    Cartwright|
|  hgusikowski|   Howard|   Gusikowski|
|    karmstrong| Kenyatta|     Armstrong|
|    drosenbaum|  Dorothy|     Rosenbaum|
|    ghalvorson|  Gaylord|     Halvorson|
|   kswaniawski|    Kyler|   Swaniawski|
|    arosenbaum|   Amanda|     Rosenbaum|
|     aschimmel| Adelbert|      Schimmel|
+-------------+----------+-------------+
only showing top 20 rows
```

[15]: 300

0.2 Casting the columns to integers

```python
[54]:   ## casting ids and ratings to integers
        from pyspark.sql.functions import col

        albums_ratings_1 = df_albums_ratings.withColumn('rating',
         ↪df_albums_ratings['rating'].cast("double").alias('album_rating'))
        albums_ratings_ = albums_ratings_1.select(col('document_id').
         ↪alias('album_id'),col('rating'),col('user_id'))


        artists_ratings_1 = df_artists_ratings.withColumn('rating',
         ↪df_artists_ratings['rating'].cast("double"))
        artists_ratings_ = artists_ratings_1.select(col('document_id').
         ↪alias('artist_id'),col('rating'),col('user_id'))
```

```
artists_1 = df_artists.withColumnRenamed('_id','artist_id')
artists_ = artists_1.withColumnRenamed('name','artist_name')
albums_  = df_albums.withColumnRenamed('_id','album_id')
users_   = df_users.withColumnRenamed('_id','user_id')
```

[55]:
```
artists_.printSchema()
artists_ratings_.printSchema()
albums_.printSchema()
albums_ratings_.printSchema()
users_.printSchema()
```

```
root
 |-- artist_id: string (nullable = true)
 |-- artist_name: string (nullable = true)

root
 |-- artist_id: string (nullable = true)
 |-- rating: double (nullable = true)
 |-- user_id: string (nullable = true)

root
 |-- album_id: string (nullable = true)
 |-- name: string (nullable = true)

root
 |-- album_id: string (nullable = true)
 |-- rating: double (nullable = true)
 |-- user_id: string (nullable = true)

root
 |-- user_id: string (nullable = true)
 |-- first_name: string (nullable = true)
 |-- last_name: string (nullable = true)
```

0.3 Creating unique integer ID's for artists and ratings

[18]:
```python
import pyspark.sql.functions as fun
from pyspark.sql.window import Window

artist_int_id = artists_.withColumn("artist_idx", fun.
 ↪monotonically_increasing_id())
new_artists = artist_int_id.withColumn("artist_idx", fun.row_number().
 ↪over(Window.orderBy("artist_idx")))

# new_artists.show()
```

```python
user_int_id = users_.withColumn("user_idx", fun.monotonically_increasing_id())
new_users = user_int_id.withColumn("user_idx", fun.row_number().over(Window.
 ↪orderBy("user_idx")))


# new_users.show()


ratings_df1 = new_artists.join(artists_ratings_,'artist_id')
# ratings_df1.show()


ratings_df2 = ratings_df1.join(new_users,'user_id')
# ratings_df2.show()


ratings_df3 = ratings_df2.drop('artist_id')
ratings_df = ratings_df3.drop('user_id')
ratings_df.show()
```

21/08/28 00:29:51 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 00:29:51 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

```
+--------------------+----------+------+----------+---------+--------+
|         artist_name|artist_idx|rating|first_name|last_name|user_idx|
+--------------------+----------+------+----------+---------+--------+
|Johann Ludwig Dei…|      4103|   1.5|   Quinton|   Kuphal|     147|
|      Wojciech Kilar|      8156|   1.0|     Grace|  Keebler|     135|
|                Me G|      1061|   3.0| Annabelle|  Kovacek|      31|
|Johann Ludwig Dei…|      4103|   4.5| Genevieve|    Koepp|     136|
|      Marianne Hoppe|       486|   3.5|     Paolo|     Roob|     295|
|            Ken Mary|      7399|   4.0|      Liza|    Hayes|     284|
|   Christopher Smith|      4577|   4.0|    Anabel|   Raynor|     156|
|   Jānis Volkinšteins|      3181|   4.0|      Maia|   Turner|     231|
|     Jacob Appelbaum|      9710|   1.5|    Alexis|     Rice|     273|
|    Demetrio Stratos|      2591|   1.0|      Hank|Greenholt|      56|
|     Laura McCormick|         7|   0.5|    Burley|  Cormier|      97|
|Robert Pete Williams|      7170|   1.5|    Alaina|    Hoppe|     212|
|       Ilmari Kianto|      7815|   3.5|   Patrick|    Sipes|     208|
|      Andreas Melzer|      1381|   5.0|Maximilian|     Beer|     177|
|       Terry Zwigoff|      5393|   3.0|    Dorcas| Nitzsche|      50|
|      Sally Liebling|      6705|   1.0|     Grant|  Denesik|     134|
|                MINX|      1417|   4.5|    Roslyn|    Frami|     235|
|         Naked Lunch|      5996|   4.5|    Mellie|  Farrell|     117|
|Ludwig van Beethoven|      1257|   0.5|    Vallie|     Bins|     288|
|                  R5|      7357|   2.5|     Elvie|  Streich|     130|
+--------------------+----------+------+----------+---------+--------+
```
only showing top 20 rows

21/08/28 00:29:52 WARN TaskSetManager: Stage 19 contains a task of very large size (15491 KiB). The maximum recommended task size is 1000 KiB.

0.4 Creating unique integer ID's for albums and ratings

```python
[56]: album_int_id = albums_.withColumn("album_idx", fun.
      monotonically_increasing_id())
      new_albums = album_int_id.withColumn("album_idx", fun.row_number().over(Window.
      orderBy("album_idx")))

      # new_albums.show()

      alb_ratings_df1 = new_albums.join(albums_ratings_,'album_id')
      # ratings_df1.show()

      alb_ratings_df2 = alb_ratings_df1.join(new_users,'user_id')
      # alb_ratings_df2.show()

      alb_ratings_df3 = alb_ratings_df2.drop('album_id')
      alb_ratings_df = alb_ratings_df3.drop('user_id')
      alb_ratings_df.show()
```

21/08/28 01:08:54 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:08:54 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:08:56 WARN TaskSetManager: Stage 4943 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:08:56 WARN TaskSetManager: Stage 4944 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
[Stage 4944:>   (0 + 2) / 2]

```
+-------------------+---------+------+----------+---------+--------+
|               name|album_idx|rating|first_name|last_name|user_idx|
+-------------------+---------+------+----------+---------+--------+
|An Evening Wasted…|      378|   0.5|   Jeramie|    Bruen|     198|
|An Evening Wasted…|      378|   1.5|   Earnest|     Wolf|     251|
|An Evening Wasted…|      378|   5.0| Stephania|  Spencer|      44|
|An Evening Wasted…|      378|   3.0|  Kenyatta|Armstrong|      15|
|An Evening Wasted…|      378|   2.0|    Richie|    Yundt|     236|
|An Evening Wasted…|      378|   1.5|     Marta|    Stark|     265|
|An Evening Wasted…|      378|   4.5|   Eleazar|     Bins|     249|
|An Evening Wasted…|      378|   2.5|       Dax|  Kovacek|     191|
|An Evening Wasted…|      378|   1.5|     Chyna| Lubowitz|      79|
```

```
|An Evening Wasted…|     378|   2.5|    Micah|    Davis|    263|
|An Evening Wasted…|     378|   4.0|   Kristy|    Smith|    225|
|An Evening Wasted…|     378|   0.5|    Elyse|   Conroy|    193|
|An Evening Wasted…|     378|   3.5|   Emilie|  Flatley|    103|
|An Evening Wasted…|     378|   2.5|Annabelle|  Kovacek|     31|
|An Evening Wasted…|     378|   4.0| Giuseppe| Nitzsche|     23|
|An Evening Wasted…|     378|   1.0|    Grant|  Denesik|    134|
|An Evening Wasted…|     378|   3.5|   Creola|   Bednar|    158|
|An Evening Wasted…|     378|   0.5|   Jovany|     Moen|    281|
|An Evening Wasted…|     378|   1.0|     Meda|    Adams|    285|
|An Evening Wasted…|     378|   4.5| Adelbert| Schimmel|     20|
+------------------+--------+------+---------+---------+-------+
only showing top 20 rows
```

[57]:
```python
ratings_df.cache()
```

21/08/28 01:09:02 WARN CacheManager: Asked to cache already cached data.

[57]: DataFrame[artist_name: string, artist_idx: int, rating: double, first_name: string, last_name: string, user_idx: int]

0.5 Counting the number of records in the data

[58]:
```python
print(ratings_df.count())
print(artists_.count())
print(users_.count())
print(albums_.count())
print(albums_ratings_.count())
```

21/08/28 01:09:19 WARN TaskSetManager: Stage 4947 contains a task of very large size (15491 KiB). The maximum recommended task size is 1000 KiB.

316065
10281

21/08/28 01:09:21 WARN TaskSetManager: Stage 4953 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.

300
6117
175413

21/08/28 01:09:21 WARN TaskSetManager: Stage 4955 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.

0.6 Finding the highest rated artists and most active users

```python
[22]: import matplotlib.pyplot as plt
      import seaborn as sns

      active_users = ratings_df.groupBy('user_idx').agg({"user_idx":"count"}).
       ↪orderBy('count(user_idx)',ascending = False)
      artists_avgrating = ratings_df.groupBy('artist_idx').agg({"rating":"avg"}).
       ↪orderBy('avg(rating)',ascending = False)
      active_users.show()
      artists_avgrating.show()

      most_active_users = active_users.select('user_idx').where(␣
       ↪active_users['count(user_idx)'] > 1950)
      highest_rated_artists = artists_avgrating.select('artist_idx').
       ↪where(artists_avgrating['avg(rating)'] > 3.49)

      active_pop_df = ratings_df.drop('artist_name','first_name','last_name')
      active_pop_df = active_pop_df.join(most_active_users,"user_idx").
       ↪join(highest_rated_artists,"artist_idx").orderBy("rating").toPandas()


      active_pop_df = active_pop_df.
       ↪pivot_table(index='artist_idx',values='rating',columns='user_idx')
      plt.figure(figsize=(12,12))
      plt.title("Ratings given by the most active users to the highest rated artists␣
       ↪")
      # sns.heatmap(active_pop_df)
```

21/08/28 00:32:30 WARN TaskSetManager: Stage 42 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:32:32 WARN TaskSetManager: Stage 44 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

```
+--------+---------------+
|user_idx|count(user_idx)|
+--------+---------------+
|     228|           2044|
|     263|           2030|
|     178|           2028|
|      83|           2025|
|     182|           2022|
|      75|           2013|
|     254|           2008|
|     291|           1992|
|     105|           1985|
|     104|           1982|
|     139|           1977|
```

```
|      37|          1976|
|      66|          1960|
|     257|          1959|
|     224|          1956|
|      98|          1955|
|     220|          1937|
|     188|          1924|
|     136|          1924|
|     287|          1920|
+--------+--------------+
only showing top 20 rows


+----------+-----------------+
|artist_idx|      avg(rating)|
+----------+-----------------+
|      6788|             3.75|
|      8089| 3.740740740740741|
|      7544|            3.725|
|      1819|3.7142857142857144|
|      4243|3.6666666666666665|
|      3136| 3.652173913043478|
|      7140| 3.642857142857143|
|      7317| 3.606060606060606|
|     10046|              3.6|
|      8918|3.5961538461538463|
|      2341| 3.586206896551724|
|      5338|3.5714285714285716|
|      8622|3.5416666666666665|
|      1037|3.5344827586206895|
|      9058|3.5166666666666666|
|      7303| 3.514705882352941|
|      6316|              3.5|
|      4708|              3.5|
|      4928|              3.5|
|     10204|              3.5|
+----------+-----------------+
only showing top 20 rows


21/08/28 00:32:33 WARN TaskSetManager: Stage 46 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:32:34 WARN TaskSetManager: Stage 48 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:32:35 WARN TaskSetManager: Stage 50 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:32:35 WARN TaskSetManager: Stage 51 contains a task of very large
```

size (15491 KiB). The maximum recommended task size is 1000 KiB.

1 Artists

1.1 Creating training validation and test sets

```
[23]: artists_train, artists_val, artists_test = ratings_df.randomSplit([0.7,0.15,0.
      ↪15])
```

```
[24]: x = artists_train.count()
      print("number of records in training set ",x)

      x = artists_val.count()
      print("number of records in training set ",x)

      x = artists_test.count()
      print("number of records in training set ",x)
```

```
21/08/28 00:32:36 WARN TaskSetManager: Stage 53 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:32:36 WARN TaskSetManager: Stage 55 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
```

```
number of records in training set  221240
number of records in training set  47638
```

```
21/08/28 00:32:37 WARN TaskSetManager: Stage 57 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
```

```
number of records in training set  47187
```

1.2 Model Training

```
[52]: from pyspark.ml.recommendation import ALS
      from pyspark.ml.evaluation import RegressionEvaluator
      import numpy as np

      reg = RegressionEvaluator(metricName="rmse", predictionCol="prediction",
       ↪labelCol="rating")
      als = ALS(maxIter=4,ratingCol="rating",userCol="user_idx", itemCol="artist_idx")

      num_ft = [2,4,8,12,16,24,30]
      reg_strength = [0.01,0.05,0.1,0.25,0.5,1,2.5]

      err_ = np.zeros((7,7))

      for i,n_ft in enumerate(num_ft):
          for j,lambda_ in enumerate(reg_strength):
              als.setParams(rank = n_ft, regParam = lambda_)
```

```
    model = als.fit(artists_train)

    predictions_1 = model.transform(artists_val)

    predictions_ = predictions_1.filter(predictions_1.prediction.
↪isNotNull())

    err_[i][j] = reg.evaluate(predictions_)

    print('Error for: number of features =  %d, regulariztion strength = %.
↪2f :: %f'% (n_ft,lambda_,err_[i][j]))
```

21/08/28 00:53:32 WARN TaskSetManager: Stage 2608 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:53:32 WARN TaskSetManager: Stage 2609 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:53:34 WARN TaskSetManager: Stage 2650 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:53:39 WARN TaskSetManager: Stage 2655 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 2, regulariztion strength = 0.01 :: 1.519245

21/08/28 00:53:40 WARN TaskSetManager: Stage 2656 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:53:41 WARN TaskSetManager: Stage 2696 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:53:47 WARN TaskSetManager: Stage 2702 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 2, regulariztion strength = 0.05 :: 1.508207

21/08/28 00:53:47 WARN TaskSetManager: Stage 2703 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:53:48 WARN TaskSetManager: Stage 2743 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:53:54 WARN TaskSetManager: Stage 2749 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 2, regulariztion strength = 0.10 :: 1.496989

21/08/28 00:53:54 WARN TaskSetManager: Stage 2750 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:53:56 WARN TaskSetManager: Stage 2791 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:01 WARN TaskSetManager: Stage 2796 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 2, regulariztion strength = 0.25 :: 1.489906

21/08/28 00:54:01 WARN TaskSetManager: Stage 2797 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:03 WARN TaskSetManager: Stage 2837 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 2, regulariztion strength = 0.50 :: 1.545485

21/08/28 00:54:09 WARN TaskSetManager: Stage 2843 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:09 WARN TaskSetManager: Stage 2844 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:10 WARN TaskSetManager: Stage 2885 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:16 WARN TaskSetManager: Stage 2890 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 2, regulariztion strength = 1.00 :: 1.756939

21/08/28 00:54:16 WARN TaskSetManager: Stage 2891 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:17 WARN TaskSetManager: Stage 2932 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:23 WARN TaskSetManager: Stage 2937 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 2, regulariztion strength = 2.50 :: 3.092774

21/08/28 00:54:23 WARN TaskSetManager: Stage 2938 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:25 WARN TaskSetManager: Stage 2979 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:30 WARN TaskSetManager: Stage 2984 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 4, regulariztion strength = 0.01 :: 1.605597

21/08/28 00:54:30 WARN TaskSetManager: Stage 2985 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:32 WARN TaskSetManager: Stage 3025 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:37 WARN TaskSetManager: Stage 3031 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 4, regulariztion strength = 0.05 :: 1.569687

21/08/28 00:54:37 WARN TaskSetManager: Stage 3032 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:39 WARN TaskSetManager: Stage 3072 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:44 WARN TaskSetManager: Stage 3078 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 4, regulariztion strength = 0.10 :: 1.533273

21/08/28 00:54:44 WARN TaskSetManager: Stage 3079 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:46 WARN TaskSetManager: Stage 3119 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:52 WARN TaskSetManager: Stage 3125 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 4, regulariztion strength = 0.25 :: 1.492465

21/08/28 00:54:52 WARN TaskSetManager: Stage 3126 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:53 WARN TaskSetManager: Stage 3167 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:54:59 WARN TaskSetManager: Stage 3172 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 4, regulariztion strength = 0.50 :: 1.541179

21/08/28 00:54:59 WARN TaskSetManager: Stage 3173 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:00 WARN TaskSetManager: Stage 3213 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:06 WARN TaskSetManager: Stage 3219 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 4, regulariztion strength = 1.00 :: 1.757700

21/08/28 00:55:06 WARN TaskSetManager: Stage 3220 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:08 WARN TaskSetManager: Stage 3260 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:14 WARN TaskSetManager: Stage 3266 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 4, regulariztion strength = 2.50 :: 3.091998

21/08/28 00:55:14 WARN TaskSetManager: Stage 3267 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:16 WARN TaskSetManager: Stage 3307 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:21 WARN TaskSetManager: Stage 3313 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 8, regulariztion strength = 0.01 :: 1.804235

21/08/28 00:55:21 WARN TaskSetManager: Stage 3314 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:23 WARN TaskSetManager: Stage 3354 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:28 WARN TaskSetManager: Stage 3360 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 8, regulariztion strength = 0.05 :: 1.703077

```
21/08/28 00:55:28 WARN TaskSetManager: Stage 3361 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:30 WARN TaskSetManager: Stage 3402 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
[Stage 3405:============================================>      (176 + 2) / 200]

Error for: number of features =  8, regulariztion strength = 0.10 :: 1.606667

21/08/28 00:55:35 WARN TaskSetManager: Stage 3407 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:35 WARN TaskSetManager: Stage 3408 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:37 WARN TaskSetManager: Stage 3448 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
[Stage 3452:=================================================> (194 + 2) / 200]

Error for: number of features =  8, regulariztion strength = 0.25 :: 1.499445

21/08/28 00:55:42 WARN TaskSetManager: Stage 3454 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:42 WARN TaskSetManager: Stage 3455 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:44 WARN TaskSetManager: Stage 3495 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:49 WARN TaskSetManager: Stage 3501 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features =  8, regulariztion strength = 0.50 :: 1.542288

21/08/28 00:55:50 WARN TaskSetManager: Stage 3502 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:51 WARN TaskSetManager: Stage 3543 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:56 WARN TaskSetManager: Stage 3548 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features =  8, regulariztion strength = 1.00 :: 1.756679

21/08/28 00:55:57 WARN TaskSetManager: Stage 3549 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:55:58 WARN TaskSetManager: Stage 3589 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features =  8, regulariztion strength = 2.50 :: 3.095070

21/08/28 00:56:04 WARN TaskSetManager: Stage 3595 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:04 WARN TaskSetManager: Stage 3596 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:06 WARN TaskSetManager: Stage 3636 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
[Stage 3640:=================================================>   (186 + 2) / 200]
```

```
Error for: number of features =  12, regulariztion strength = 0.01 :: 2.009675

21/08/28 00:56:11 WARN TaskSetManager: Stage 3642 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:11 WARN TaskSetManager: Stage 3643 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:13 WARN TaskSetManager: Stage 3683 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:18 WARN TaskSetManager: Stage 3689 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features =  12, regulariztion strength = 0.05 :: 1.817498

21/08/28 00:56:18 WARN TaskSetManager: Stage 3690 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:20 WARN TaskSetManager: Stage 3730 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:25 WARN TaskSetManager: Stage 3736 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features =  12, regulariztion strength = 0.10 :: 1.654361

21/08/28 00:56:25 WARN TaskSetManager: Stage 3737 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:27 WARN TaskSetManager: Stage 3777 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:33 WARN TaskSetManager: Stage 3783 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features =  12, regulariztion strength = 0.25 :: 1.506819

21/08/28 00:56:33 WARN TaskSetManager: Stage 3784 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:35 WARN TaskSetManager: Stage 3824 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:40 WARN TaskSetManager: Stage 3830 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features =  12, regulariztion strength = 0.50 :: 1.539239

21/08/28 00:56:40 WARN TaskSetManager: Stage 3831 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:42 WARN TaskSetManager: Stage 3872 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:47 WARN TaskSetManager: Stage 3877 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features =  12, regulariztion strength = 1.00 :: 1.759108

21/08/28 00:56:47 WARN TaskSetManager: Stage 3878 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:49 WARN TaskSetManager: Stage 3918 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
```

21/08/28 00:56:55 WARN TaskSetManager: Stage 3924 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 12, regulariztion strength = 2.50 :: 3.087291

21/08/28 00:56:55 WARN TaskSetManager: Stage 3925 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:56:57 WARN TaskSetManager: Stage 3966 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:03 WARN TaskSetManager: Stage 3971 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 16, regulariztion strength = 0.01 :: 2.204849

21/08/28 00:57:03 WARN TaskSetManager: Stage 3972 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:05 WARN TaskSetManager: Stage 4013 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 16, regulariztion strength = 0.05 :: 1.909432

21/08/28 00:57:11 WARN TaskSetManager: Stage 4018 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:11 WARN TaskSetManager: Stage 4019 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:13 WARN TaskSetManager: Stage 4059 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:18 WARN TaskSetManager: Stage 4065 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 16, regulariztion strength = 0.10 :: 1.685028

21/08/28 00:57:18 WARN TaskSetManager: Stage 4066 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:20 WARN TaskSetManager: Stage 4107 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:25 WARN TaskSetManager: Stage 4112 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 16, regulariztion strength = 0.25 :: 1.508764

21/08/28 00:57:26 WARN TaskSetManager: Stage 4113 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:27 WARN TaskSetManager: Stage 4154 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:33 WARN TaskSetManager: Stage 4159 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 16, regulariztion strength = 0.50 :: 1.538842

21/08/28 00:57:33 WARN TaskSetManager: Stage 4160 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:35 WARN TaskSetManager: Stage 4200 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

21/08/28 00:57:40 WARN TaskSetManager: Stage 4206 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 16, regulariztion strength = 1.00 :: 1.758975

21/08/28 00:57:40 WARN TaskSetManager: Stage 4207 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:42 WARN TaskSetManager: Stage 4248 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:48 WARN TaskSetManager: Stage 4253 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 16, regulariztion strength = 2.50 :: 3.088942

21/08/28 00:57:48 WARN TaskSetManager: Stage 4254 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:51 WARN TaskSetManager: Stage 4294 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:56 WARN TaskSetManager: Stage 4300 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 24, regulariztion strength = 0.01 :: 2.583199

21/08/28 00:57:56 WARN TaskSetManager: Stage 4301 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:57:58 WARN TaskSetManager: Stage 4342 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:58:04 WARN TaskSetManager: Stage 4347 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 24, regulariztion strength = 0.05 :: 1.984289

21/08/28 00:58:04 WARN TaskSetManager: Stage 4348 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:58:07 WARN TaskSetManager: Stage 4389 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
[Stage 4392:===> (194 + 2) / 200]

Error for: number of features = 24, regulariztion strength = 0.10 :: 1.695115

21/08/28 00:58:13 WARN TaskSetManager: Stage 4394 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:58:13 WARN TaskSetManager: Stage 4395 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:58:15 WARN TaskSetManager: Stage 4436 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:58:20 WARN TaskSetManager: Stage 4441 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 24, regulariztion strength = 0.25 :: 1.513990

21/08/28 00:58:21 WARN TaskSetManager: Stage 4442 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:58:23 WARN TaskSetManager: Stage 4483 contains a task of very large

size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:58:28 WARN TaskSetManager: Stage 4488 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 24, regulariztion strength = 0.50 :: 1.539897

21/08/28 00:58:28 WARN TaskSetManager: Stage 4489 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:58:30 WARN TaskSetManager: Stage 4529 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:58:36 WARN TaskSetManager: Stage 4535 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 24, regulariztion strength = 1.00 :: 1.759179

21/08/28 00:58:36 WARN TaskSetManager: Stage 4536 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:58:38 WARN TaskSetManager: Stage 4576 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:58:44 WARN TaskSetManager: Stage 4582 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 24, regulariztion strength = 2.50 :: 3.087301

21/08/28 00:58:44 WARN TaskSetManager: Stage 4583 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:58:46 WARN TaskSetManager: Stage 4624 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:58:52 WARN TaskSetManager: Stage 4629 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 30, regulariztion strength = 0.01 :: 2.780608

21/08/28 00:58:52 WARN TaskSetManager: Stage 4630 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:58:55 WARN TaskSetManager: Stage 4671 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
[Stage 4674:===>   (172 + 2) / 200]

Error for: number of features = 30, regulariztion strength = 0.05 :: 1.986181

21/08/28 00:59:00 WARN TaskSetManager: Stage 4676 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:59:00 WARN TaskSetManager: Stage 4677 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:59:03 WARN TaskSetManager: Stage 4718 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:59:08 WARN TaskSetManager: Stage 4723 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 30, regulariztion strength = 0.10 :: 1.677161

21/08/28 00:59:08 WARN TaskSetManager: Stage 4724 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

```
21/08/28 00:59:11 WARN TaskSetManager: Stage 4764 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:59:16 WARN TaskSetManager: Stage 4770 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features =  30, regulariztion strength = 0.25 :: 1.512113

21/08/28 00:59:16 WARN TaskSetManager: Stage 4771 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:59:18 WARN TaskSetManager: Stage 4811 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:59:25 WARN TaskSetManager: Stage 4817 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features =  30, regulariztion strength = 0.50 :: 1.540100

21/08/28 00:59:25 WARN TaskSetManager: Stage 4818 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:59:27 WARN TaskSetManager: Stage 4858 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:59:33 WARN TaskSetManager: Stage 4864 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features =  30, regulariztion strength = 1.00 :: 1.759207

21/08/28 00:59:33 WARN TaskSetManager: Stage 4865 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 00:59:36 WARN TaskSetManager: Stage 4906 contains a task of very large
size (15491 KiB). The maximum recommended task size is 1000 KiB.
[Stage 4909:======================================>                  (152 + 2) / 200]

Error for: number of features =  30, regulariztion strength = 2.50 :: 3.086808
```

1.3 Error Matrix and Optimal Paramaters

```python
[53]: print(err_) ## the errors for different hyperparameters
opt_reg = 0.25
opt_ft = 4
reg_idx = 3
ft_idx = 1
als_opt = als.setParams(regParam = reg_strength[reg_idx], rank = num_ft[ft_idx])
model = als_opt.fit(ratings_df)
```

```
[[1.51924471 1.50820654 1.49698915 1.48990564 1.54548478 1.75693913
  3.09277425]
 [1.60559713 1.56968668 1.5332726  1.49246515 1.54117895 1.75769986
  3.09199837]
 [1.80423523 1.70307659 1.60666693 1.49944457 1.5422884  1.75667933
  3.09507011]
 [2.00967452 1.81749754 1.65436097 1.50681903 1.53923902 1.75910829
```

```
  3.08729099]
 [2.2048488  1.90943164 1.68502839 1.50876392 1.53884172 1.75897527
  3.08894189]
 [2.58319885 1.98428904 1.69511474 1.51398979 1.53989743 1.75917899
  3.08730122]
 [2.78060802 1.98618114 1.67716072 1.51211327 1.54010041 1.75920712
  3.08680788]]
```

21/08/28 01:00:34 WARN TaskSetManager: Stage 4911 contains a task of very large size (15491 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:00:34 WARN TaskSetManager: Stage 4912 contains a task of very large size (15491 KiB). The maximum recommended task size is 1000 KiB.

1.4 Top 5 artists for all users

[29]:
```
userRecs = model.recommendForAllUsers(5)
userRecs.orderBy('user_idx').show(n=10,truncate = False)
```

```
[Stage 2406:====================================>                  (137 + 2) / 200]

+--------+--------------------------------------------------------------------
--------------------------+
|user_idx|recommendations
|
+--------+--------------------------------------------------------------------
--------------------------+
|1       |[{7544, 3.4648542}, {8089, 3.4069488}, {8918, 3.389954}, {4243,
3.374133}, {1819, 3.3651853}]    |
|2       |[{1819, 3.4734209}, {6788, 3.4438586}, {10046, 3.4365757}, {8089,
3.4164448}, {3136, 3.4134703}]|
|3       |[{7544, 3.572998}, {8918, 3.5232759}, {8089, 3.4610038}, {4243,
3.429198}, {7303, 3.4211671}]    |
|4       |[{4243, 3.460855}, {7544, 3.4414763}, {8089, 3.4352136}, {6788,
3.4141283}, {1819, 3.412143}]    |
|5       |[{1819, 3.4749312}, {6788, 3.4640107}, {8089, 3.4639509}, {7544,
3.4615295}, {7140, 3.3880632}] |
|6       |[{1819, 3.4821591}, {6788, 3.4166138}, {8089, 3.4141784}, {10046,
3.412616}, {3136, 3.3967154}] |
|7       |[{1819, 3.451817}, {6788, 3.3686082}, {8089, 3.3638687}, {7396,
3.3417702}, {7140, 3.316318}]    |
|8       |[{1819, 3.307413}, {7140, 3.2820246}, {3281, 3.2656436}, {7396,
3.241385}, {9058, 3.2397218}]    |
|9       |[{1819, 3.5749135}, {6788, 3.5678859}, {8089, 3.563372}, {7544,
3.5494065}, {4243, 3.521646}]    |
|10      |[{6788, 3.3855686}, {1819, 3.3791702}, {8089, 3.3229852}, {3136,
3.2989779}, {10046, 3.2980933}]|
|11      |[{6788, 3.5100856}, {8089, 3.3529837}, {7544, 3.3292365}, {1819,
3.3289943}, {4243, 3.327856}]   |
|12      |[{4243, 3.466939}, {7544, 3.4622812}, {8089, 3.4159293}, {6788,
```

```
3.409471}, {3136, 3.372313}]      |
|13      |[{7544, 3.321055}, {8089, 3.2933087}, {7694, 3.289783}, {1819,
3.2871528}, {6788, 3.281153}]      |
|14      |[{6788, 3.4836507}, {8089, 3.4181724}, {1819, 3.4081907}, {7544,
3.4033754}, {4243, 3.3972945}] |
|15      |[{6788, 3.7318196}, {7544, 3.6090531}, {8089, 3.605756}, {1819,
3.565579}, {4243, 3.557609}]      |
|16      |[{7544, 3.5039945}, {8089, 3.437168}, {8918, 3.4336588}, {7140,
3.3966186}, {1819, 3.3965044}]  |
|17      |[{1819, 3.4780521}, {8089, 3.46336}, {7544, 3.4569733}, {6788,
3.4369898}, {4243, 3.4125302}]    |
|18      |[{1819, 3.409272}, {8089, 3.380494}, {7544, 3.3656988}, {4243,
3.3528113}, {7317, 3.349691}]      |
|19      |[{6788, 3.4364657}, {7544, 3.4156857}, {8089, 3.4005911}, {4243,
3.3934436}, {1819, 3.3613064}] |
|20      |[{4243, 3.4773178}, {7544, 3.470709}, {8089, 3.4408097}, {3136,
3.4128938}, {7317, 3.405713}]    |
+--------+-------------------------------------------------------------------
------------------------+
only showing top 20 rows
```

1.5 Top 5 users for all artists

```
[31]: artistRecs = model.recommendForAllItems(5)
      artistRecs.orderBy('artist_idx').show(truncate = False)
```

```
[Stage 2432:=================================================>  (188 + 2) / 200]

+----------+--------------------------------------------------------------------
---------------------+
|artist_idx|recommendations
|
+----------+--------------------------------------------------------------------
---------------------+
|1         |[{64, 2.884092}, {167, 2.8634398}, {189, 2.8480952}, {143,
2.8216105}, {272, 2.8073788}]   |
|2         |[{264, 2.7811399}, {189, 2.714577}, {144, 2.66073}, {71, 2.5928388},
{30, 2.5895445}]       |
|3         |[{189, 2.697942}, {264, 2.526031}, {144, 2.524995}, {131,
2.5183878}, {143, 2.4762745}]     |
|4         |[{64, 2.9404852}, {167, 2.923203}, {189, 2.9186113}, {143,
2.8987393}, {272, 2.851272}]     |
|5         |[{189, 3.1480453}, {264, 3.0771575}, {144, 3.0213897}, {131,
2.972484}, {71, 2.938232}]     |
|6         |[{264, 3.763568}, {144, 3.3275576}, {30, 3.241963}, {151,
3.2176673}, {71, 3.2032735}]      |
```

24

```
|7         |[{189, 3.1044228}, {143, 2.8891954}, {131, 2.8633952}, {64,
2.8444715}, {43, 2.8209248}]   |
|8         |[{264, 3.5237117}, {144, 3.0860772}, {143, 3.06461}, {64, 3.030457},
{151, 3.0045516}]      |
|9         |[{143, 3.0613923}, {64, 2.9367957}, {167, 2.8408816}, {294,
2.7931166}, {269, 2.765064}]   |
|10        |[{264, 3.4908469}, {151, 3.1962962}, {144, 3.1732826}, {189,
3.1644626}, {143, 3.152787}] |
|11        |[{189, 2.5658092}, {264, 2.5379772}, {143, 2.4576797}, {144,
2.4489255}, {151, 2.4389088}]|
|12        |[{143, 3.2258387}, {64, 3.1485415}, {167, 3.0361145}, {264,
2.9771037}, {189, 2.9705563}] |
|13        |[{143, 2.851377}, {189, 2.817823}, {64, 2.8137417}, {264,
2.7377827}, {167, 2.7138677}]    |
|14        |[{189, 3.1127043}, {64, 3.0391598}, {167, 3.0114129}, {272,
2.9813826}, {143, 2.975492}]  |
|15        |[{264, 2.6821935}, {189, 2.4896502}, {143, 2.4782264}, {144,
2.4730852}, {64, 2.4636908}] |
|16        |[{189, 3.3240476}, {143, 3.3008766}, {64, 3.295181}, {264,
3.2745268}, {167, 3.1791563}]   |
|17        |[{189, 2.4175165}, {131, 2.2400663}, {80, 2.2350767}, {64,
2.2104702}, {214, 2.1906555}]   |
|18        |[{264, 2.8485634}, {64, 2.7145987}, {143, 2.672307}, {249,
2.634594}, {180, 2.6139188}]    |
|19        |[{264, 3.621532}, {143, 3.2480478}, {144, 3.2114146}, {151,
3.18923}, {189, 3.1394775}]    |
|20        |[{264, 3.296989}, {144, 2.7776203}, {30, 2.711888}, {143,
2.6899176}, {28, 2.6811585}]     |
+----------+--------------------------------------------------------------------
---------------------+
only showing top 20 rows
```

2 Albums

2.1 Creating training validation and test sets

```python
albums_train, albums_val, albums_test = alb_ratings_df.randomSplit([0.7,0.15,0.
 →15])
```

```python
x = albums_train.count()
print("number of records in training set ",x)

x = albums_val.count()
print("number of records in training set ",x)
```

```
x = albums_test.count()
print("number of records in training set ",x)
```

21/08/28 01:18:18 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:18:18 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:18:19 WARN TaskSetManager: Stage 5027 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:18:20 WARN TaskSetManager: Stage 5029 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:18:21 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:18:21 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

number of records in training set 123004

21/08/28 01:18:22 WARN TaskSetManager: Stage 5034 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:18:22 WARN TaskSetManager: Stage 5036 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:18:24 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:18:24 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

number of records in training set 26407

21/08/28 01:18:25 WARN TaskSetManager: Stage 5042 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:18:25 WARN TaskSetManager: Stage 5041 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
[Stage 5044:===============================>   (119 + 2) / 200]

number of records in training set 26002
```

## 2.2 Model Training

```python
[66]: from pyspark.ml.recommendation import ALS
      from pyspark.ml.evaluation import RegressionEvaluator
      import numpy as np

      reg = RegressionEvaluator(metricName="rmse", predictionCol="prediction",
       ↪labelCol="rating")
      als = ALS(maxIter=4,ratingCol="rating",userCol="user_idx", itemCol="album_idx")

      num_ft = [2,4,8,12,16,24,30]
      reg_strength = [0.01,0.05,0.1,0.25,0.5,1,2.5]

      err_ = np.zeros((7,7))

      for i,n_ft in enumerate(num_ft):
          for j,lambda_ in enumerate(reg_strength):
              als.setParams(rank = n_ft, regParam = lambda_)

              model = als.fit(albums_train)

              predictions_1 = model.transform(albums_val)

              predictions_ = predictions_1.filter(predictions_1.prediction.
       ↪isNotNull())

              err_[i][j] = reg.evaluate(predictions_)

              print('Error for: number of features =  %d, regulariztion strength = %.
       ↪2f :: %f'% (n_ft,lambda_,err_[i][j]))
```

21/08/28 01:19:51 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:19:51 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:19:52 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:19:52 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:19:53 WARN TaskSetManager: Stage 5050 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:19:53 WARN TaskSetManager: Stage 5051 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.

21/08/28 01:19:59 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:19:59 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:20:00 WARN TaskSetManager: Stage 5111 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:20:01 WARN TaskSetManager: Stage 5113 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:20:10 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:20:10 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 2, regulariztion strength = 0.01 :: 1.526608

21/08/28 01:20:11 WARN TaskSetManager: Stage 5122 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:20:11 WARN TaskSetManager: Stage 5124 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:20:17 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:20:17 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:20:18 WARN TaskSetManager: Stage 5184 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:20:19 WARN TaskSetManager: Stage 5183 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 2, regulariztion strength = 0.05 :: 1.514184

21/08/28 01:20:27 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:20:27 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:20:28 WARN TaskSetManager: Stage 5194 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:20:28 WARN TaskSetManager: Stage 5195 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:20:34 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:20:34 WARN WindowExec: No Partition Defined for Window operation!

Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:20:36 WARN TaskSetManager: Stage 5255 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:20:36 WARN TaskSetManager: Stage 5256 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
[Stage 5262:===> (191 + 2) / 200]

Error for: number of features = 2, regulariztion strength = 0.10 :: 1.501820

21/08/28 01:20:45 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:20:45 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:20:46 WARN TaskSetManager: Stage 5266 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:20:46 WARN TaskSetManager: Stage 5267 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:20:52 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:20:52 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:20:53 WARN TaskSetManager: Stage 5326 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:20:54 WARN TaskSetManager: Stage 5328 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:21:02 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:21:02 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance degradation.

Error for: number of features = 2, regulariztion strength = 0.25 :: 1.492091

21/08/28 01:21:04 WARN TaskSetManager: Stage 5338 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:21:04 WARN TaskSetManager: Stage 5339 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:21:10 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:21:10 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:21:11 WARN TaskSetManager: Stage 5399 contains a task of very large

size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:21:11 WARN TaskSetManager: Stage 5401 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:21:20 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:21:20 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 2, regulariztion strength = 0.50 :: 1.546688

21/08/28 01:21:21 WARN TaskSetManager: Stage 5410 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:21:21 WARN TaskSetManager: Stage 5411 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:21:27 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:21:27 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:21:28 WARN TaskSetManager: Stage 5472 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:21:28 WARN TaskSetManager: Stage 5471 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:21:37 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:21:37 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 2, regulariztion strength = 1.00 :: 1.764298

21/08/28 01:21:38 WARN TaskSetManager: Stage 5482 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:21:38 WARN TaskSetManager: Stage 5483 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:21:44 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:21:44 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:21:45 WARN TaskSetManager: Stage 5542 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:21:45 WARN TaskSetManager: Stage 5544 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:21:54 WARN WindowExec: No Partition Defined for Window operation!

Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:21:54 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 2, regulariztion strength = 2.50 :: 3.099820

21/08/28 01:21:55 WARN TaskSetManager: Stage 5554 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:21:55 WARN TaskSetManager: Stage 5556 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:22:02 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:22:02 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:22:03 WARN TaskSetManager: Stage 5614 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:22:03 WARN TaskSetManager: Stage 5615 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:22:11 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:22:11 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 4, regulariztion strength = 0.01 :: 1.631999

21/08/28 01:22:12 WARN TaskSetManager: Stage 5626 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:22:12 WARN TaskSetManager: Stage 5628 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:22:18 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:22:18 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:22:19 WARN TaskSetManager: Stage 5687 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:22:19 WARN TaskSetManager: Stage 5688 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:22:28 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:22:28 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance

degradation.

Error for: number of features = 4, regulariztion strength = 0.05 :: 1.588694

21/08/28 01:22:29 WARN TaskSetManager: Stage 5698 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:22:29 WARN TaskSetManager: Stage 5700 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:22:35 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:22:35 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:22:36 WARN TaskSetManager: Stage 5745 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:22:36 WARN TaskSetManager: Stage 5747 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:22:45 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:22:45 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 4, regulariztion strength = 0.10 :: 1.546728

21/08/28 01:22:46 WARN TaskSetManager: Stage 5770 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:22:46 WARN TaskSetManager: Stage 5772 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:22:51 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:22:51 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:22:53 WARN TaskSetManager: Stage 5817 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:22:53 WARN TaskSetManager: Stage 5819 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:23:01 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:23:01 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 4, regulariztion strength = 0.25 :: 1.497212

21/08/28 01:23:03 WARN TaskSetManager: Stage 5842 contains a task of very large

size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:23:03 WARN TaskSetManager: Stage 5844 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:23:09 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:23:09 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:23:10 WARN TaskSetManager: Stage 5889 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:23:11 WARN TaskSetManager: Stage 5891 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 4, regulariztion strength = 0.50 :: 1.543218

21/08/28 01:23:21 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:23:21 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:23:22 WARN TaskSetManager: Stage 5914 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:23:23 WARN TaskSetManager: Stage 5916 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:23:29 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:23:29 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:23:30 WARN TaskSetManager: Stage 5975 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:23:30 WARN TaskSetManager: Stage 5977 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:23:38 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:23:38 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.

Error for: number of features = 4, regulariztion strength = 1.00 :: 1.765313

21/08/28 01:23:39 WARN TaskSetManager: Stage 5986 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:23:40 WARN TaskSetManager: Stage 5987 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:23:45 WARN WindowExec: No Partition Defined for Window operation!

Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:23:45 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:23:46 WARN TaskSetManager: Stage 6033 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:23:46 WARN TaskSetManager: Stage 6034 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 4, regulariztion strength = 2.50 :: 3.097599

21/08/28 01:23:55 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:23:55 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:23:56 WARN TaskSetManager: Stage 6058 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:23:56 WARN TaskSetManager: Stage 6060 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:24:02 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:24:02 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:24:03 WARN TaskSetManager: Stage 6120 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:24:03 WARN TaskSetManager: Stage 6119 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:24:12 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:24:12 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 8, regulariztion strength = 0.01 :: 1.845785

21/08/28 01:24:13 WARN TaskSetManager: Stage 6130 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:24:13 WARN TaskSetManager: Stage 6131 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:24:19 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:24:19 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance

degradation.
21/08/28 01:24:20 WARN TaskSetManager: Stage 6191 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:24:21 WARN TaskSetManager: Stage 6192 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:24:29 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:24:29 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 8, regulariztion strength = 0.05 :: 1.729940

21/08/28 01:24:30 WARN TaskSetManager: Stage 6202 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:24:30 WARN TaskSetManager: Stage 6204 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:24:36 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:24:36 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:24:37 WARN TaskSetManager: Stage 6249 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:24:38 WARN TaskSetManager: Stage 6251 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
[Stage 6270:===>   (177 + 3) / 200]

Error for: number of features = 8, regulariztion strength = 0.10 :: 1.623955

21/08/28 01:24:46 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:24:46 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:24:47 WARN TaskSetManager: Stage 6274 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:24:47 WARN TaskSetManager: Stage 6276 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:24:53 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:24:53 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:24:55 WARN TaskSetManager: Stage 6334 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.

21/08/28 01:24:55 WARN TaskSetManager: Stage 6335 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 8, regulariztion strength = 0.25 :: 1.506741

21/08/28 01:25:03 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:25:03 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:25:05 WARN TaskSetManager: Stage 6346 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:25:05 WARN TaskSetManager: Stage 6348 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:25:11 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:25:11 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:25:12 WARN TaskSetManager: Stage 6406 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:25:12 WARN TaskSetManager: Stage 6408 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:25:21 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:25:21 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 8, regulariztion strength = 0.50 :: 1.542578

21/08/28 01:25:22 WARN TaskSetManager: Stage 6418 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:25:23 WARN TaskSetManager: Stage 6420 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:25:28 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:25:28 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:25:29 WARN TaskSetManager: Stage 6479 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:25:30 WARN TaskSetManager: Stage 6480 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:25:38 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance

degradation.
21/08/28 01:25:38 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 8, regulariztion strength = 1.00 :: 1.764978

21/08/28 01:25:39 WARN TaskSetManager: Stage 6490 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:25:39 WARN TaskSetManager: Stage 6492 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:25:45 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:25:45 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:25:46 WARN TaskSetManager: Stage 6537 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:25:46 WARN TaskSetManager: Stage 6538 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:25:54 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:25:54 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 8, regulariztion strength = 2.50 :: 3.100518

21/08/28 01:25:55 WARN TaskSetManager: Stage 6562 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:25:56 WARN TaskSetManager: Stage 6563 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:26:02 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:26:02 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:26:03 WARN TaskSetManager: Stage 6623 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:26:03 WARN TaskSetManager: Stage 6625 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:26:11 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:26:11 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 12, regulariztion strength = 0.01 :: 2.081808

21/08/28 01:26:12 WARN TaskSetManager: Stage 6634 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:26:12 WARN TaskSetManager: Stage 6636 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:26:18 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:26:18 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:26:19 WARN TaskSetManager: Stage 6681 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:26:19 WARN TaskSetManager: Stage 6683 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
[Stage 6702:===>   (188 + 2) / 200]

Error for: number of features = 12, regulariztion strength = 0.05 :: 1.864750

21/08/28 01:26:27 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:26:27 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:26:28 WARN TaskSetManager: Stage 6706 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:26:29 WARN TaskSetManager: Stage 6707 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:26:34 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:26:34 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:26:35 WARN TaskSetManager: Stage 6767 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:26:35 WARN TaskSetManager: Stage 6769 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:26:44 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:26:44 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 12, regulariztion strength = 0.10 :: 1.682877

21/08/28 01:26:45 WARN TaskSetManager: Stage 6778 contains a task of very large

size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:26:45 WARN TaskSetManager: Stage 6780 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:26:50 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:26:50 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:26:51 WARN TaskSetManager: Stage 6839 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:26:52 WARN TaskSetManager: Stage 6840 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:26:59 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:26:59 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.

Error for: number of features = 12, regulariztion strength = 0.25 :: 1.515756

21/08/28 01:27:01 WARN TaskSetManager: Stage 6850 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:01 WARN TaskSetManager: Stage 6851 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:06 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:27:06 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:27:07 WARN TaskSetManager: Stage 6911 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:07 WARN TaskSetManager: Stage 6913 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:16 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:27:16 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.

Error for: number of features = 12, regulariztion strength = 0.50 :: 1.543491

21/08/28 01:27:17 WARN TaskSetManager: Stage 6922 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:17 WARN TaskSetManager: Stage 6924 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:23 WARN WindowExec: No Partition Defined for Window operation!

Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:27:23 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:27:24 WARN TaskSetManager: Stage 6969 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:24 WARN TaskSetManager: Stage 6970 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:33 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:27:33 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 12, regulariztion strength = 1.00 :: 1.765974

21/08/28 01:27:34 WARN TaskSetManager: Stage 6994 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:34 WARN TaskSetManager: Stage 6995 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:40 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:27:40 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:27:42 WARN TaskSetManager: Stage 7041 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:42 WARN TaskSetManager: Stage 7043 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:50 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:27:50 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 12, regulariztion strength = 2.50 :: 3.093008

21/08/28 01:27:51 WARN TaskSetManager: Stage 7066 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:52 WARN TaskSetManager: Stage 7068 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:57 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:27:57 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance

degradation.
21/08/28 01:27:59 WARN TaskSetManager: Stage 7113 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:27:59 WARN TaskSetManager: Stage 7114 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:28:07 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:28:07 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 16, regulariztion strength = 0.01 :: 2.289665

21/08/28 01:28:08 WARN TaskSetManager: Stage 7138 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:28:09 WARN TaskSetManager: Stage 7139 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:28:14 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:28:14 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:28:16 WARN TaskSetManager: Stage 7200 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:28:16 WARN TaskSetManager: Stage 7199 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:28:25 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:28:25 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 16, regulariztion strength = 0.05 :: 1.945005

21/08/28 01:28:26 WARN TaskSetManager: Stage 7210 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:28:27 WARN TaskSetManager: Stage 7212 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:28:32 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:28:32 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:28:34 WARN TaskSetManager: Stage 7257 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:28:34 WARN TaskSetManager: Stage 7258 contains a task of very large

size (8651 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 16, regulariztion strength = 0.10 :: 1.705709

21/08/28 01:28:42 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:28:42 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:28:44 WARN TaskSetManager: Stage 7282 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:28:44 WARN TaskSetManager: Stage 7284 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:28:50 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:28:50 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:28:51 WARN TaskSetManager: Stage 7344 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:28:51 WARN TaskSetManager: Stage 7343 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 16, regulariztion strength = 0.25 :: 1.517224

21/08/28 01:29:00 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:29:00 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:29:01 WARN TaskSetManager: Stage 7354 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:29:01 WARN TaskSetManager: Stage 7355 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:29:08 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:29:08 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:29:09 WARN TaskSetManager: Stage 7415 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:29:09 WARN TaskSetManager: Stage 7417 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:29:17 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

21/08/28 01:29:17 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 16, regulariztion strength = 0.50 :: 1.542779

21/08/28 01:29:19 WARN TaskSetManager: Stage 7426 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:29:19 WARN TaskSetManager: Stage 7428 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:29:26 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:29:26 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:29:27 WARN TaskSetManager: Stage 7487 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:29:27 WARN TaskSetManager: Stage 7488 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:29:36 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:29:36 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 16, regulariztion strength = 1.00 :: 1.765875

21/08/28 01:29:37 WARN TaskSetManager: Stage 7498 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:29:37 WARN TaskSetManager: Stage 7500 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:29:44 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:29:44 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:29:45 WARN TaskSetManager: Stage 7545 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:29:45 WARN TaskSetManager: Stage 7546 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:29:54 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:29:54 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 16, regulariztion strength = 2.50 :: 3.095969

21/08/28 01:29:55 WARN TaskSetManager: Stage 7570 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:29:55 WARN TaskSetManager: Stage 7571 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:30:02 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:30:02 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:30:03 WARN TaskSetManager: Stage 7617 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:30:04 WARN TaskSetManager: Stage 7618 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
[Stage 7638:===>   (176 + 2) / 200]

Error for: number of features = 24, regulariztion strength = 0.01 :: 2.670063

21/08/28 01:30:12 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:30:12 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:30:13 WARN TaskSetManager: Stage 7642 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:30:14 WARN TaskSetManager: Stage 7643 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:30:20 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:30:20 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:30:21 WARN TaskSetManager: Stage 7689 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:30:21 WARN TaskSetManager: Stage 7690 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:30:29 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:30:29 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 24, regulariztion strength = 0.05 :: 1.990277

21/08/28 01:30:30 WARN TaskSetManager: Stage 7714 contains a task of very large

size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:30:31 WARN TaskSetManager: Stage 7715 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:30:37 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:30:37 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:30:38 WARN TaskSetManager: Stage 7774 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:30:38 WARN TaskSetManager: Stage 7776 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 24, regulariztion strength = 0.10 :: 1.696234

21/08/28 01:30:46 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:30:46 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:30:47 WARN TaskSetManager: Stage 7786 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:30:47 WARN TaskSetManager: Stage 7787 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:30:53 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:30:53 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:30:55 WARN TaskSetManager: Stage 7848 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:30:55 WARN TaskSetManager: Stage 7847 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 24, regulariztion strength = 0.25 :: 1.520737

21/08/28 01:31:03 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:31:03 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:31:04 WARN TaskSetManager: Stage 7858 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:31:05 WARN TaskSetManager: Stage 7860 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:31:11 WARN WindowExec: No Partition Defined for Window operation!

Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:31:11 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:31:12 WARN TaskSetManager: Stage 7905 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:31:12 WARN TaskSetManager: Stage 7906 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:31:20 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:31:20 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.

Error for: number of features = 24, regulariztion strength = 0.50 :: 1.544622

21/08/28 01:31:21 WARN TaskSetManager: Stage 7930 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:31:22 WARN TaskSetManager: Stage 7931 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:31:28 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:31:28 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:31:29 WARN TaskSetManager: Stage 7990 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:31:30 WARN TaskSetManager: Stage 7991 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 24, regulariztion strength = 1.00 :: 1.766001

21/08/28 01:31:38 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:31:38 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:31:39 WARN TaskSetManager: Stage 8002 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:31:40 WARN TaskSetManager: Stage 8003 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:31:46 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:31:46 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance

degradation.
21/08/28 01:31:47 WARN TaskSetManager: Stage 8064 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:31:47 WARN TaskSetManager: Stage 8063 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 24, regulariztion strength = 2.50 :: 3.092488

21/08/28 01:31:55 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:31:55 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:31:57 WARN TaskSetManager: Stage 8074 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:31:57 WARN TaskSetManager: Stage 8075 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:32:03 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:32:03 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:32:05 WARN TaskSetManager: Stage 8134 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:32:05 WARN TaskSetManager: Stage 8136 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:32:13 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:32:13 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 30, regulariztion strength = 0.01 :: 2.894250

21/08/28 01:32:14 WARN TaskSetManager: Stage 8146 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:32:14 WARN TaskSetManager: Stage 8147 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:32:21 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:32:21 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:32:22 WARN TaskSetManager: Stage 8193 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:32:22 WARN TaskSetManager: Stage 8195 contains a task of very large

size (8651 KiB). The maximum recommended task size is 1000 KiB.
[Stage 8214:==> (178 + 2) / 200]

Error for: number of features = 30, regulariztion strength = 0.05 :: 2.025447

21/08/28 01:32:31 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:32:31 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:32:32 WARN TaskSetManager: Stage 8218 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:32:32 WARN TaskSetManager: Stage 8219 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:32:38 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:32:38 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:32:39 WARN TaskSetManager: Stage 8279 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:32:39 WARN TaskSetManager: Stage 8281 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 30, regulariztion strength = 0.10 :: 1.701827

21/08/28 01:32:48 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:32:48 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:32:49 WARN TaskSetManager: Stage 8290 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:32:49 WARN TaskSetManager: Stage 8291 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:32:55 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:32:55 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:32:57 WARN TaskSetManager: Stage 8351 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:32:57 WARN TaskSetManager: Stage 8353 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 30, regulariztion strength = 0.25 :: 1.524825

21/08/28 01:33:06 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:33:06 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:33:08 WARN TaskSetManager: Stage 8362 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:33:08 WARN TaskSetManager: Stage 8363 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:33:14 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:33:14 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:33:15 WARN TaskSetManager: Stage 8424 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:33:16 WARN TaskSetManager: Stage 8423 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:33:24 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:33:24 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.

Error for: number of features = 30, regulariztion strength = 0.50 :: 1.544528

21/08/28 01:33:25 WARN TaskSetManager: Stage 8434 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:33:25 WARN TaskSetManager: Stage 8436 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:33:32 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:33:32 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:33:33 WARN TaskSetManager: Stage 8495 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:33:33 WARN TaskSetManager: Stage 8497 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.

Error for: number of features = 30, regulariztion strength = 1.00 :: 1.765995

21/08/28 01:33:43 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:33:43 WARN WindowExec: No Partition Defined for Window operation!

Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:33:44 WARN TaskSetManager: Stage 8506 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:33:44 WARN TaskSetManager: Stage 8507 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:33:50 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:33:50 WARN WindowExec: No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation.
21/08/28 01:33:51 WARN TaskSetManager: Stage 8566 contains a task of very large size (7450 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:33:51 WARN TaskSetManager: Stage 8568 contains a task of very large size (8651 KiB). The maximum recommended task size is 1000 KiB.
[Stage 8574:==>   (170 + 2) / 200]

Error for: number of features = 30, regulariztion strength = 2.50 :: 3.093556

2.3   Error Matrix and Optimal Paramaters

```python
print(err_) ## the errors for different hyperparameters
opt_reg = 0.25
opt_ft = 4
reg_idx = 3
ft_idx = 1
als_opt = als.setParams(regParam = reg_strength[reg_idx], rank = num_ft[ft_idx])
model = als_opt.fit(alb_ratings_df)
```

```
[[1.52660774 1.51418443 1.50181993 1.49209092 1.54668753 1.76429842
  3.09982001]
 [1.63199909 1.58869437 1.5467275  1.49721221 1.54321751 1.76531283
  3.0975987 ]
 [1.84578507 1.72994036 1.62395505 1.50674053 1.54257751 1.76497766
  3.10051762]
 [2.08180761 1.86475041 1.68287715 1.5157565  1.54349092 1.76597432
  3.09300834]
 [2.28966507 1.94500527 1.70570945 1.51722393 1.54277946 1.7658755
  3.09596892]
 [2.67006306 1.99027667 1.69623448 1.52073693 1.54462243 1.76600061
  3.09248772]
 [2.89424951 2.0254473  1.70182685 1.52482462 1.54452769 1.76599528
  3.09355614]]
```

21/08/28 01:34:17 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance

degradation.
21/08/28 01:34:17 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:34:18 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:34:18 WARN WindowExec: No Partition Defined for Window operation!
Moving all data to a single partition, this can cause serious performance
degradation.
21/08/28 01:34:19 WARN TaskSetManager: Stage 8580 contains a task of very large
size (8651 KiB). The maximum recommended task size is 1000 KiB.
21/08/28 01:34:19 WARN TaskSetManager: Stage 8581 contains a task of very large
size (7450 KiB). The maximum recommended task size is 1000 KiB.

2.4 Top 5 albums for all users

```
[68]: userRecs = model.recommendForAllUsers(5)
      userRecs.orderBy('user_idx').show(truncate = False)
```

```
[Stage 8640:=================================>                    (133 + 2) / 200]

+--------+-----------------------------------------------------------------
------------------------+
|user_idx|recommendations
|
+--------+-----------------------------------------------------------------
------------------------+
|1       |[{3552, 3.4766335}, {5662, 3.4522738}, {4441, 3.3930807}, {3576,
3.3058467}, {546, 3.282145}]   |
|2       |[{3552, 3.7208512}, {2572, 3.4706404}, {1799, 3.3244174}, {5662,
3.3223758}, {4480, 3.316272}] |
|3       |[{3552, 3.4937115}, {5662, 3.3197756}, {229, 3.2247553}, {1522,
3.2152987}, {4480, 3.180952}]   |
|4       |[{3552, 3.6456823}, {5662, 3.3895822}, {229, 3.3698626}, {4491,
3.3664784}, {1522, 3.3604074}] |
|5       |[{3552, 3.6469088}, {5662, 3.4438152}, {4441, 3.4265494}, {1548,
3.3235533}, {1382, 3.3176494}]|
|6       |[{3552, 3.6184654}, {2572, 3.3113327}, {4441, 3.2970376}, {4435,
3.258345}, {1548, 3.2432368}] |
|7       |[{3552, 3.824387}, {1382, 3.785201}, {4441, 3.7541711}, {4018,
3.7470376}, {1495, 3.7282395}]   |
|8       |[{3552, 3.7381964}, {2572, 3.5504458}, {4480, 3.4063163}, {501,
3.386284}, {2789, 3.334941}]    |
|9       |[{3552, 3.6470704}, {5662, 3.3300326}, {2572, 3.28463}, {1799,
3.275526}, {1548, 3.2710557}]    |
|10      |[{3552, 3.6892438}, {4018, 3.354984}, {2572, 3.3415124}, {4261,
3.313704}, {1495, 3.3128772}]   |
```

```
|11      |[{4441, 3.665763}, {3552, 3.5442967}, {1127, 3.5015936}, {3860,
3.4173522}, {5662, 3.3979075}] |
|12      |[{3552, 3.567127}, {5662, 3.4071715}, {229, 3.3473172}, {1522,
3.3402543}, {5333, 3.3103113}]   |
|13      |[{3552, 3.6670837}, {4441, 3.5063539}, {5662, 3.493185}, {4480,
3.4290495}, {1127, 3.3735294}] |
|14      |[{3552, 3.7056987}, {5662, 3.5585682}, {229, 3.4604504}, {1522,
3.4520266}, {4480, 3.3939943}] |
|15      |[{3552, 3.562554}, {5662, 3.363091}, {1382, 3.284894}, {1495,
3.2786744}, {4441, 3.2631855}]    |
|16      |[{3552, 3.6410174}, {4441, 3.4430418}, {5662, 3.4359245}, {4480,
3.357992}, {1548, 3.3185163}] |
|17      |[{3552, 3.738447}, {2572, 3.4199002}, {5662, 3.3953695}, {4441,
3.3627248}, {1548, 3.354663}]   |
|18      |[{3552, 3.5945516}, {5662, 3.4072886}, {4480, 3.3286974}, {4441,
3.2887397}, {229, 3.285813}]   |
|19      |[{5662, 3.5696313}, {3552, 3.5689688}, {229, 3.444121}, {1522,
3.4400396}, {3797, 3.4367657}]   |
|20      |[{3552, 3.615436}, {5662, 3.3589065}, {229, 3.3182864}, {1522,
3.307881}, {2572, 3.304473}]     |
+--------+----------------------------------------------------------------
-----------------------+
only showing top 20 rows
```

2.5 Top 5 users for all albums

```
albumsRecs = model.recommendForAllItems(5)
albumsRecs.orderBy('album_idx').show(truncate = False)
```

```
[Stage 8656:============================================>      (174 + 2) / 200]

+--------+----------------------------------------------------------------
------------------+
|album_idx|recommendations
|
+--------+----------------------------------------------------------------
------------------+
|1       |[{82, 2.5937362}, {244, 2.3745542}, {32, 2.3609848}, {164,
2.3591018}, {130, 2.3095777}]|
|2       |[{82, 2.8947434}, {244, 2.8884087}, {7, 2.8417134}, {290, 2.830026},
{32, 2.8233843}]    |
|3       |[{82, 3.2091103}, {32, 2.9694686}, {33, 2.9693875}, {244, 2.9483056},
{7, 2.934648}]    |
|4       |[{164, 3.1925972}, {82, 2.9778757}, {33, 2.9481614}, {195,
2.9478085}, {11, 2.846305}]   |
|5       |[{82, 2.9472923}, {33, 2.8257322}, {164, 2.816277}, {195, 2.7796202},
```

```
{99, 2.7455535}]  |
|6        |[{82, 2.7199583}, {244, 2.5240297}, {293, 2.5201957}, {139,
2.3802133}, {68, 2.3492107}]|
|7        |[{82, 3.3330982}, {32, 3.1026945}, {244, 3.090713}, {130, 3.0240521},
{66, 3.0154753}]  |
|8        |[{244, 2.7068338}, {82, 2.6342711}, {290, 2.6103976}, {7, 2.6015778},
{32, 2.5527716}]  |
|9        |[{82, 3.1599185}, {293, 2.905383}, {164, 2.8331532}, {260, 2.765773},
{83, 2.762621}]   |
|10       |[{82, 3.2181177}, {293, 2.992}, {244, 2.9236805}, {260, 2.8232043},
{139, 2.7982724}]   |
|11       |[{33, 3.0181737}, {290, 2.9116879}, {7, 2.8824906}, {82, 2.8009672},
{99, 2.7829456}]   |
|12       |[{82, 3.1135523}, {293, 2.977402}, {33, 2.9181266}, {244, 2.884265},
{99, 2.8330088}]   |
|13       |[{244, 3.1548333}, {82, 3.138816}, {293, 3.0388196}, {7, 2.8626027},
{29, 2.8622887}]   |
|14       |[{82, 2.7590942}, {244, 2.515597}, {293, 2.4687443}, {164, 2.468445},
{32, 2.4560676}]  |
|15       |[{33, 3.011405}, {99, 2.7777536}, {293, 2.7668405}, {290, 2.7647467},
{82, 2.761252}]   |
|16       |[{293, 3.0317855}, {82, 2.9707735}, {244, 2.875142}, {160, 2.777089},
{139, 2.7755919}] |
|17       |[{33, 2.9756532}, {82, 2.9272234}, {99, 2.8134313}, {293, 2.7995434},
{7, 2.771349}]    |
|18       |[{33, 2.819974}, {7, 2.817862}, {290, 2.8098307}, {82, 2.7669802},
{32, 2.7265267}]     |
|19       |[{82, 3.0123923}, {293, 2.8457513}, {139, 2.7273612}, {244,
2.709247}, {68, 2.5836885}] |
|20       |[{164, 3.3207622}, {195, 2.9149442}, {90, 2.7913604}, {83,
2.7274146}, {82, 2.7139816}] |
+---------+-----------------------------------------------------------------
------------------+
only showing top 20 rows
```

```
[33]:  # from pymongo import MongoClient
```

```
[34]:  # client = MongoClient("mongodb://localhost:8888")
```

```
[35]:  # mydb = client["mydatabase"]
```

```
[36]:  # artists = mydb["artists"]
       # temp = [{'a':'sdf','b':'asdf'}]
       # artists.insert_many(temp)
```

```
---------------------------------------------------------------------------
ServerSelectionTimeoutError                Traceback (most recent call last)
/tmp/ipykernel_69/2286772032.py in <module>
      1 artists = mydb["artists"]
      2 temp = [{'a':'sdf','b':'asdf'}]
----> 3 artists.insert_many(temp)

/opt/conda/lib/python3.9/site-packages/pymongo/collection.py in␣
 ↪insert_many(self, documents, ordered, bypass_document_validation, session)
    768            blk = _Bulk(self, ordered, bypass_document_validation)
    769            blk.ops = [doc for doc in gen()]
--> 770            blk.execute(write_concern, session=session)
    771            return InsertManyResult(inserted_ids, write_concern.acknowledge)
    772

/opt/conda/lib/python3.9/site-packages/pymongo/bulk.py in execute(self,␣
 ↪write_concern, session)
    527                self.execute_no_results(sock_info, generator)
    528            else:
--> 529                return self.execute_command(generator, write_concern,␣
 ↪session)
    530
    531

/opt/conda/lib/python3.9/site-packages/pymongo/bulk.py in execute_command(self,␣
 ↪generator, write_concern, session)
    357
    358            client = self.collection.database.client
--> 359            with client._tmp_session(session) as s:
    360                client._retry_with_session(
    361                    self.is_retryable, retryable_bulk, s, self)

/opt/conda/lib/python3.9/contextlib.py in __enter__(self)
    115        del self.args, self.kwds, self.func
    116        try:
--> 117            return next(self.gen)
    118        except StopIteration:
    119            raise RuntimeError("generator didn't yield") from None

/opt/conda/lib/python3.9/site-packages/pymongo/mongo_client.py in␣
 ↪_tmp_session(self, session, close)
   1946                return
   1947
-> 1948            s = self._ensure_session(session)
   1949            if s:
   1950                try:
```

```
/opt/conda/lib/python3.9/site-packages/pymongo/mongo_client.py in
↪_ensure_session(self, session)
   1933              # Don't make implicit sessions causally consistent.
↪Applications
   1934              # should always opt-in.
-> 1935              return self.__start_session(True, causal_consistency=False)
   1936         except (ConfigurationError, InvalidOperation):
   1937              # Sessions not supported, or multiple users authenticated.


/opt/conda/lib/python3.9/site-packages/pymongo/mongo_client.py in
↪__start_session(self, implicit, **kwargs)
   1881
   1882         # Raises ConfigurationError if sessions are not supported.
-> 1883         server_session = self._get_server_session()
   1884         opts = client_session.SessionOptions(**kwargs)
   1885         return client_session.ClientSession(


/opt/conda/lib/python3.9/site-packages/pymongo/mongo_client.py in
↪_get_server_session(self)
   1919     def _get_server_session(self):
   1920         """Internal: start or resume a _ServerSession."""
-> 1921         return self._topology.get_server_session()
   1922
   1923     def _return_server_session(self, server_session, lock):


/opt/conda/lib/python3.9/site-packages/pymongo/topology.py in
↪get_server_session(self)
    518              # Sessions are always supported in load balanced mode.
    519              if not self._settings.load_balanced:
--> 520                  session_timeout = self._check_session_support()
    521              else:
    522                  # Sessions never time out in load balanced mode.


/opt/conda/lib/python3.9/site-packages/pymongo/topology.py in
↪_check_session_support(self)
    497             if self._description.topology_type == TOPOLOGY_TYPE.Single:
    498                 if not self._description.has_known_servers:
--> 499                     self._select_servers_loop(
    500                         any_server_selector,
    501                         self._settings.server_selection_timeout,


/opt/conda/lib/python3.9/site-packages/pymongo/topology.py in
↪_select_servers_loop(self, selector, timeout, address)
    216              # No suitable servers.
    217              if timeout == 0 or now > end_time:
--> 218                  raise ServerSelectionTimeoutError(

    219                      "%s, Timeout: %ss, Topology Description: %r" %
```

```
    220                             (self._error_message(selector), timeout, self.
 ↪description))

ServerSelectionTimeoutError: localhost:8888: timed out, Timeout: 30s, Topology␣
 ↪Description: <TopologyDescription id: 612985b2d5cd04a3a0853d40, topology_type ␣
 ↪Single, servers: [<ServerDescription ('localhost', 8888) server_type: Unknown ␣
 ↪rtt: None, error=NetworkTimeout('localhost:8888: timed out')>]>
```

```python
[ ]: # with open('dataset-full-ratings/artists/all_artists.json') as file_names:
     #     for file_name in file_names:
     #         with open(file_name) as f:
     #             file_data = json.load(f)  # load data from JSON to dict
     #             for k, v in file_data.items():  # iterate over key-value pairs
     #                 collection.insert_one(v)  # your collection object here
```

```
[ ]:
```