CS5542 Big Data Apps and Analytics

LAB ASSIGNMENT #3

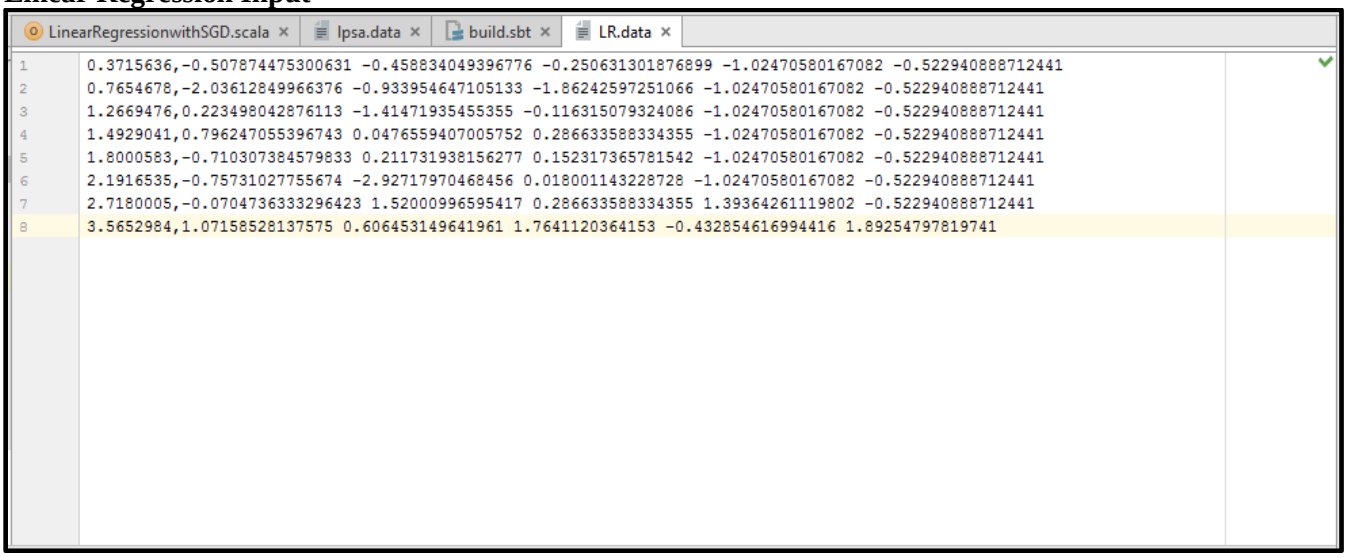
Name:Sidrah Junaid
Class Id:15

1. **Spark Programming:**

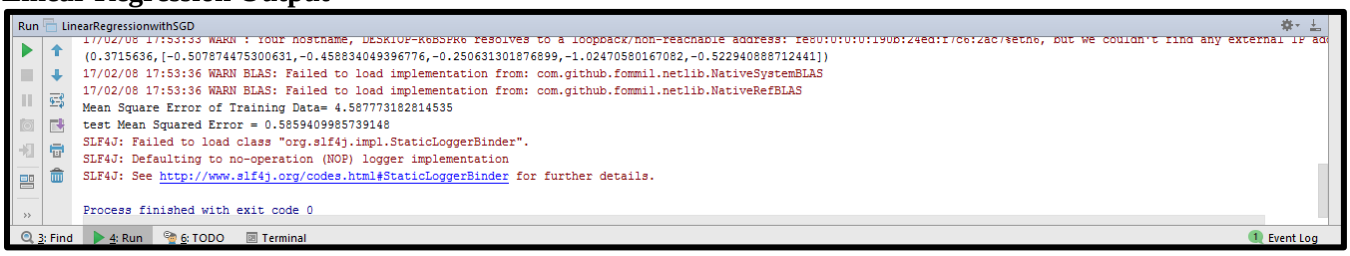
Write a spark program for the following Machine Learning Tasks.

1. Implement to build a linear regression model for selected parameters for chimpanzee's daily movement, activities and interaction. Define your own datasets.
   Here the selected parameters are the sleeping time ,eating habits grooming and fighting parameters against each chimpanzee.

Linear Regression Input



Linear Regression Output



2. Implement K-Means clustering for the clusters of the chimpanzee's activities. Define your own data sets.
   Same dataset used for K-Mean Clustering

K-Mean Clustering Input

build.sbt × | kMeansClustering.scala × | kmeans_data.txt × | KMean.txt ×

```
0.3715636 -0.507874475300631 -0.458834049396776 -0.250631301876899 -1.02470580167082 -0.522940888712441
0.7654678 -2.03612849966376 -0.933954647105133 -1.86242597251066 -1.02470580167082 -0.522940888712441
1.2669476 0.223498042876113 -1.41471935455355 -0.116315079324086 -1.02470580167082 -0.522940888712441
1.4929041 0.796247055396743 0.0476559407005752 0.286633588334355 -1.02470580167082 -0.522940888712441
1.8000583 -0.710307384579833 0.211731938156277 0.152317365781542 -1.02470580167082 -0.522940888712441
2.1916535 -0.75731027755674 -2.92717970468456 0.018001143228728 -1.02470580167082 -0.522940888712441
2.7180005 -0.0704736333296423 1.52000996595417 0.286633588334355 1.39364261119802 -0.522940888712441
3.5652984 1.07158528137575 0.606453149641961 1.7641120364153 -0.432854616994416 1.89254797819741
```

K-Mean Clustering Output

Run kMeansClustering

```
17/02/08 18:07:52 WARN : Your hostname, DESKTOP-K6BSPR6 resolves to a loopback/non-reachable address: fe80:0:0:0:190b:24ed:f7c6:2ac7%eth6, but we couldn't find any external IP ad
[1.8000583,-0.710307384579833,0.211731938156277,0.152317365781542,-1.02470580167082,-0.522940888712441]
[0.3715636,-0.507874475300631,-0.458834049396776,-0.250631301876899,-1.02470580167082,-0.522940888712441]
[2.1916535,-0.75731027755674,-2.92717970468456,0.018001143228728,-1.02470580167082,-0.522940888712441]
[0.7654678,-2.03612849966376,-0.933954647105133,-1.86242597251066,-1.02470580167082,-0.522940888712441]
[2.7180005,-0.0704736333296423,1.52000996595417,0.286633588334355,1.39364261119802,-0.522940888712441]
[1.2669476,0.223498042876113,-1.41471935455355,-0.116315079324086,-1.02470580167082,-0.522940888712441]
[3.5652984,1.07158528137575,0.606453149641961,1.7641120364153,-0.432854616994416,1.89254797819741]
[1.4929041,0.796247055396743,0.0476559407005752,0.286633588334355,-1.02470580167082,-0.522940888712441]
17/02/08 18:07:55 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeSystemBLAS
17/02/08 18:07:55 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeRefBLAS
Within Set Sum of Squared Errors = 36.445794583365796
Clustering on training data:
([1.8000583,-0.710307384579833,0.211731938156277,0.152317365781542,-1.02470580167082,-0.522940888712441],0)
([2.1916535,-0.75731027755674,-2.92717970468456,0.018001143228728,-1.02470580167082,-0.522940888712441],1)
([2.7180005,-0.0704736333296423,1.52000996595417,0.286633588334355,1.39364261119802,-0.522940888712441],0)
([3.5652984,1.07158528137575,0.606453149641961,1.7641120364153,-0.432854616994416,1.89254797819741],0)
([0.3715636,-0.507874475300631,-0.458834049396776,-0.250631301876899,-1.02470580167082,-0.522940888712441],0)
([0.7654678,-2.03612849966376,-0.933954647105133,-1.86242597251066,-1.02470580167082,-0.522940888712441],0)
([1.2669476,0.223498042876113,-1.41471935455355,-0.116315079324086,-1.02470580167082,-0.522940888712441],0)
([1.4929041,0.796247055396743,0.0476559407005752,0.286633588334355,-1.02470580167082,-0.522940888712441],0)
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.

Process finished with exit code 0
```

2Video Annotation:

First the video sample1.mkv is provided is provided into the program and then the key frame is detected using KeyFrameDetection.java and then the frames is annotated by using imageannotation.java and then the summary was collected in text file.

Key Frames:

Image Annotation