**JAVA**

Affinchè il programma operi normalmente è necessario preparare l’ambiente di esecuzione java.

Come prima cosa scaricare ed installare Eclipse e java development kit (jdk) ed aggiungere JAVA\_HOME alle variabili di ambiente.

Per utilizzare il .jar senza librerie Spark assicurarsi che i dati del file <Nome\_file\_dati\_input.txt> abbiano come prime colonne le coordinate dei punti e come ultima colonna l'attributo CLASSE.

Assicurarsi che il file .jar e le cartelle "dati\_input" e "dati\_output" siano nella stessa Directory.

Progetto

│

│

└── BVQ.jar

│

│

└── 📂dati\_input

│ │

│ └ comando\_di\_lancio.txt

│ │

│ └ <Nome\_file\_dati\_input.txt>

│

│

└── 📂dati\_output

│

└ results.txt

Per eseguire il .jar avviare da terminale il seguente comando di lancio:

java -jar <Path>\<Nome\_file.jar>

Altrimenti è possibile trascinare il file nella finestra del terminale.

**SPARK**

Per importare le librerie di Spark modificare Eclipse in modo da supportare Maven seguendo le istruzioni nei link di seguito

Maven configuration: <https://www.youtube.com/watch?v=4sO-VgqHLp4&t=134s>

Spark configuration: <https://spark.apache.org/docs/latest/configuration.html>

A questo punto importare il progetto come Progetto Maven inserendo il nome della cartella contenente il progetto stesso come Artifact Id.

Assicurarsi che il comando di lancio all'interno del relativo file con il seguente ordine di attributi:

"

file risultati: results

replicabile: false

excel: false

numero vettori codice per classe: 5

delta: 0.2

learning rate: 0.01

percentuale train: 80

matrice dei costi: [0 1;1 0]

seed: 11

epoche BVQ: 1

flag iterazioni per bvq: true

numero bag(iterazioni BVQ): 1

dimensione bag: 10000

flag utilizzo la som: true

epoche SOM: 5

"

Assicurarsi che il dataset sia scritto nel seguente modo (seguire il modello LibSVM):

CLASSE 1:x 2:y …

e che le cartelle "dati\_input" e "dati\_output" siano nella cartella conf di hadoop-Spark.

Per creare il file .jar esportare il progetto come Runnable Jar da Eclipse.

Per avviare la prova in locale eseguire da terminale il comando:

spark-submit --class "MainClass" --master local[\*] --driver-memory 12g <Path>\<Nome\_file.jar>

Per eseguire il codice da remoto o per ulteriori informazioni riguardo gli argomenti da inserire nel terminale seguire il link di seguito:

<https://spark.apache.org/docs/latest/submitting-applications.html>

Per eseguire il codice da Eclipse la prima volta andare in

Run Configuration -> MainClass -> Argument -> Program Argument

scrivere “ eclipse “, cliccare Apply e di seguito Run.

Per compilare il jar con maven:

mvn clean compile assembly:single