



Università
degli Studi
di Ferrara

Advanced School in Artificial Intelligence

Corso di Perfezionamento
Dipartimento di Ingegneria

Esercitazione pratica in Weka

Elisabetta Gentili, Damiano Azzolini

elisabetta.gentili1@unife.it, damiano.azzolini@unife.it

Anno Accademico 2021-22

Come fare Machine Learning/Data Mining

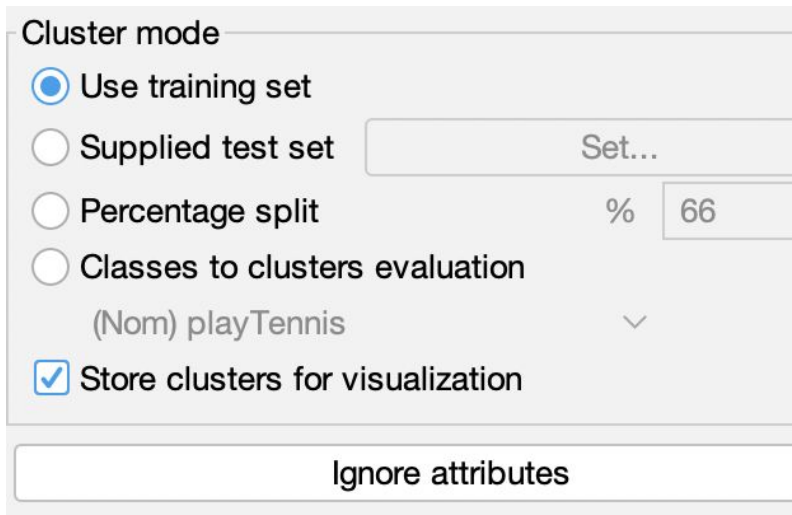


Come fare Machine Learning/Data Mining

1. Caricare i dati nell'Explorer (diversi formati possibili)
2. Selezionare l'algoritmo di apprendimento
 - pannello *Classify* per Decision trees e Random Forest
 - pannello *Cluster* per clustering
 - pannello *Associate* per regole associative
3. Valutare le performance (possibile per modelli predittivi)

Clustering con K-means

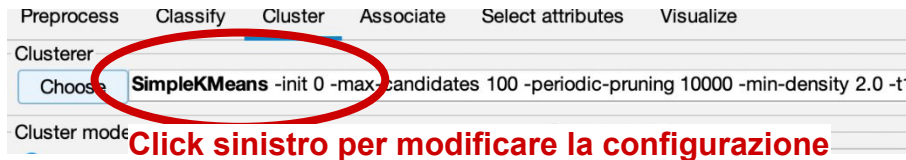
1. Carichiamo il dataset nell'Explorer
2. Selezioniamo l'algoritmo **Simple K-means** dal pannello *Cluster* e impostiamo i parametri (es. *numCluster*)



The screenshot shows a 'Cluster mode' configuration panel. It contains several radio buttons for selecting the data source: 'Use training set' (selected), 'Supplied test set' (with a 'Set...' button), 'Percentage split' (with a '%' sign and a '66' input field), and 'Classes to clusters evaluation'. Below these is a dropdown menu currently showing '(Nom) playTennis'. At the bottom, there is a checked checkbox for 'Store clusters for visualization' and an 'Ignore attributes' button.

3. Nella finestra *cluster mode* selezioniamo *use training set*.
4. Avviamo l'algoritmo
5. Valutiamo il risultato con:
Visualize Cluster Assignments
nel pannello *Visualize*

Clustering con K-means



Click sinistro per modificare la configurazione

weka.clusterers.SimpleKMeans

canopyMaxNumCanopiesToHoldInMemory	100
canopyMinimumCanopyDensity	2.0
canopyPeriodicPruningRate	10000
canopyT1	-1.25
canopyT2	-1.0
debug	False
displayStdDevs	False
distanceFunction	Choose EuclideanDistance -f
doNotCheckCapabilities	False
dontReplaceMissingValues	False
fastDistanceCalc	False
initializationMethod	Random
maxIterations	500
numClusters	2
numExecutionSlots	1
preserveInstancesOrder	False
reduceNumberOfDistanceCalcsViaCanopies	False
seed	10

Association Rules

car	False	▼
classIndex	-1	
delta	0.05	
doNotCheckCapabilities	False	▼
lowerBoundMinSupport	0.1	
metricType	Confidence	▼
minMetric	0.9	
numRules	10	
outputItemSets	False	▼
removeAllMissingCols	False	▼
significanceLevel	-1.0	
treatZeroAsMissing	False	▼
upperBoundMinSupport	1.0	
verbose	False	▼

1. Carichiamo il dataset nell'Explorer
2. Nel pannello *Associate* selezioniamo l'algoritmo **Apriori**
3. Nella finestra degli attributi configurare l'algoritmo
4. Avviamo l'algoritmo
5. Nel pannello principale saranno visualizzate le regole associative generate

Esercizio su Clustering

1. Utilizzare il dataset *iris*: il file comprende 3 classi di 50 istanze ciascuna, dove ogni classe si riferisce a una specie differente di iris.
2. Numero di istanze: 150 ognuna delle quali con 4 attributi + la classe = 5
3. Applicare l'algoritmo *SimpleKmeans*

Per il clustering bisogna utilizzare il file con o senza classe? Perché?

Esercizio su regole associative

1. Aprire in Weka il dataset `supermarket.arff`, che specifica la tipologia di prodotti acquistati insieme in 4627 diverse transazioni (spese)
2. Applicare l'algoritmo Apriori per estrarre al massimo 10 regole associative con supporto minimo = 15% e confidenza minima = 90%

Che correlazione tra prodotti viene estratta da gran parte delle regole?

Esercizio su regole associative

3. Cosa accade portando la soglia di supporto al 30%?
4. Provare ad abbassare la confidenza mantenendo `min_sup=30%`