

ESERCITAZIONE UNSUPERVISED LEARNING

Corso di Data Analytics

10 Novembre 2021

Si consideri il dataset Wine (presente sulla pagina di Virtuale del corso):

1. Usando la tecnica del KMeans (`n_clusters=3`, `init='k-means++'`, `max_iter=300`, `n_init=10`, `random_state=0`), stampare il valore di `inertia_` risultante.
2. Ripetere l'esercizio applicando una tecnica di pre-processing sui dati (es. standardizzazione/scaling).
3. Individuare il numero minimo di cluster da individuare mediante la tecnica elbow.
4. Applicare la tecnica di GaussianMixture (`n_components=3`). Costruire tre scatterplot:
 - a. `X=Alchol`, `Y=Hue`, `Colore=Class`
 - b. `X=Alchol`, `Y=Hue`, `Colore=Numero del cluster` secondo algoritmo GaussianMixture
 - c. `X=Alchol`, `Y=Hue`, `Colore= Numero del cluster` secondo algoritmo KMeans
5. Applicare la tecnica dell' AgglomerativeClustering(`compute_full_tree=True`, `affinity='euclidean'`, `linkage='ward'`), stampare a video il *dendogramma* risultante. Individuare un possibile cutpoint se si vogliono generare 5 cluster.