Classe:

Nome e Cognome:

L’algoritmo k−NN

Si ha a disposizione il seguente **dataset di training** con riportati i dati relativi alla playlist di un utente.

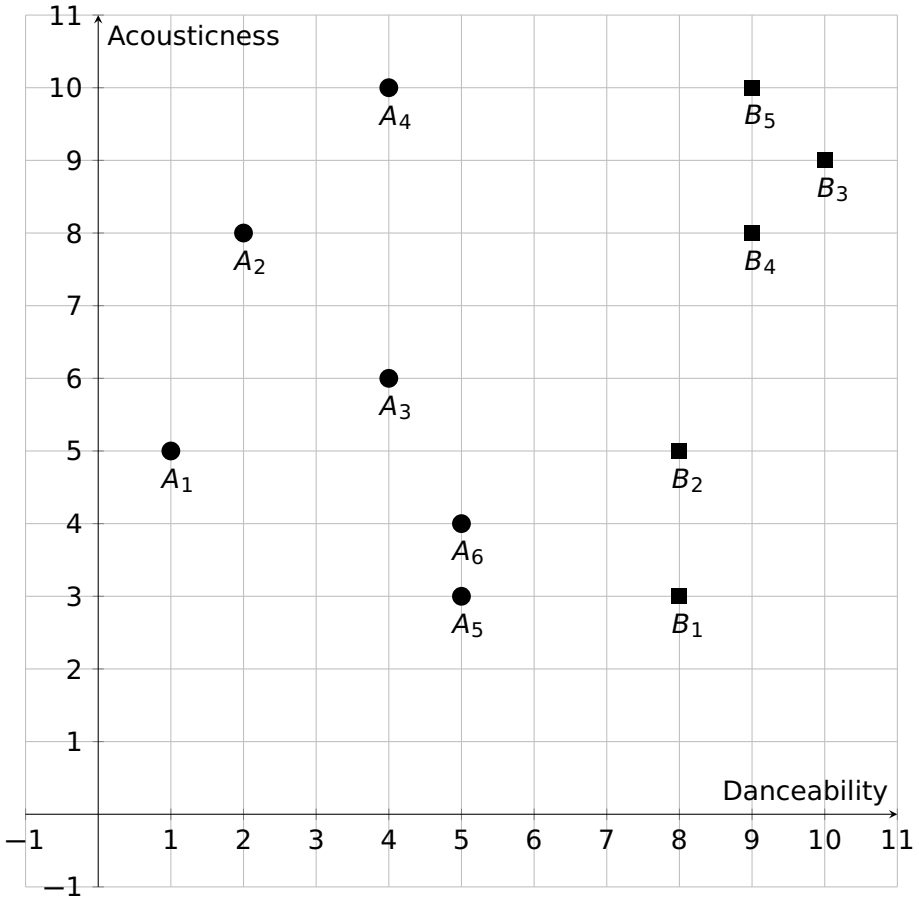
Danceability	Acousticness	Label
1	5	Like
2	8	Like
4	6	Like
4	10	Like
5	3	Like
5	4	Like
8	3	Don’t like
8	5	Don’t like
10	9	Don’t like
9	8	Don’t like
9	10	Don’t like

La prima *feature* (colonna) rappresenta la danceability di una canzone, la seconda l’acousticness. Le prime due colonne costituiscono la mia matrice X_{train}. Alll’utente della tua piattaforma solo alcune canzoni piacciono. L’informazione è riportata nella terza colonna, *label*. Questa colonna è il vettore y_{train} che voglio provare a predirre in base ai dati della matrice X_{train}.

Si ha poi a disposizione un secondo dataset su cui testare il modello, **dataset di test**.

Danceability	Acousticness	Label
5	8	Don’t like
7	4	Like
10	4	Don’t like
10	8	Don’t like
3	7	Like

1. Consideriamo nel **dataset di training** le due *feature* come le coordinate (x, y) di un punto nel piano cartesiano come riportato nel seguente grafico.



Classifica i seguenti punti $P_1(5;8)$, $P_2(7,4)$, $P_3(10,4)$, $P_4(10,8)$, $P_5(3,7)$ considerando $k = 3$.

(a) Come si calcola la distanza tra due punti non allineati orizzontalmente o verticalmente?

.....

(b) E nel caso di punti allineati orizzontalmente o verticalmente?

.....

(c) Riporta le distanze nella seguente tabella

	A_1	A_2	A_3	A_4	A_5	A_6	B_1	B_2	B_3	B_4	B_5
P_1											
P_2											
P_3											
P_4											
P_5											

(d) Ordinare le distanze in ordine crescente, calcolare la frequenza per ogni label e classificare il punto.

	Distanze ordinate			f_{Like}	$f_{\text{Don't like}}$	label
P_1						
P_2						
P_3						
P_4						
P_5						

2. Sai che i cinque punti che hai classificato precedentemente in realtà sono rispettivamente Don't like, Like, Don't like, Don't like, Like.

   (a) Completare la seguente matrice di confusione

		Valori predetti	
Valori reali			

   (b) Come si calcola l'accuratezza?

   .....

   (c) In questo caso l'accuratezza risulta .....

3. Ponendo k=5

   (a) il punto P_2 come viene classificato? Riportare i conti svolti.

   (b) Come cambia la matrice di confusione? E l'accuratezza?

		Valori predetti	
Valori reali			

   In questo caso l'accuratezza risulta .....