

# Intelligenza Artificiale

Anno accademico  
2008-2009

*Esercizi* Classificazione e Predizione

# Domanda di riscaldamento ;)

- Dato il cluster clacolato utilizzando l'algoritmo K-means standard formato dai seguenti vettori:
  - $A1=\{4,5,5,0\}$
  - $A2=\{0,2,3,5\}$
  - $A3=\{2,0,5,4\}$
  - $A4=\{5,2,0,1\}$
  - $A5=\{4,1,2,5\}$
- Qual è il centroide del cluster?

# Esercizio

## k-nn

X	Y	Class
-1	1	-
0	1	+
0	2	-
1	-1	-
1	0	+
1	2	+
2	2	-
2	3	+

- Considerare i dati di train nello spazio euclideo bidimensionale rappresentati in tabella. X e Y sono gli attributi, "Class" è la classe.
- Qual è la classe predetta da un classificatore 3-nn per l'esempio (1,1)?
- per k=5?
- per k=7?

# Esercizio naïve bayes

<b>a</b>	<b>b</b>	<b>c</b>	<b>K</b>
1	0	1	1
1	1	1	1
0	1	1	0
1	1	0	0
1	0	1	0
0	0	0	1
0	0	0	1
0	0	1	0

- Considerare i dati di train specificati da quattro attributi booleani (a,b,c,K) in cui K rappresenta la classe. Si supponga di utilizzare un classificatore Naive Bayes per determinare il valore di K a partire dagli attributi a,b,c.
- Calcolare la classe di appartenenza dell'istanza con a=1, b=1, c=0.
- Calcolare la classe di appartenenza dell'istanza con a=1, b=1 e col valore di c sconosciuto.

# Esercizio

- La vostra azienda ha circa 2000000 di clienti. Nel database aziendale, ogni cliente è descritto da una tupla con 400 attributi. Avete bisogno di un modello dei vostri clienti che spendono molto. Per questo motivo contattate un'azienda specializzata in applicazioni di data mining.,
- Vi propongono tre soluzioni: una basata su alberi di decisione, una basata su k-nn, e una basata su classificatori Naive Bayes.
- Sapendo che il modello dovrà successivamente essere implementato su un pocket pc di limitate capacità di memoria e CPU (16Mb di memoria per i dati), quali delle tre soluzioni scegliereste? Perché?

# Esercizio k-means

osservazione	X	Y
P1	1,0	1,5
P2	1,0	4,5
P3	2,0	1,5
P4	2,0	3,5
P5	3,0	2,5
P6	5,0	6,0

- Si considerino 6 punti sul piano cartesiano mostrati in tabella.
- Si simuli in modo grafico l'esecuzione dell'algoritmo K-means con  $K=2$  a partire dai centroidi iniziali P1 e P3, evidenziando i centroidi e i cluster ad ognuna delle iterazioni.

# PRISM

- Si consideri il seguente training set relativo alle promozioni per una carta di credito.
- Applicare l'algoritmo PRISM con variabile target **Promozione Assicurazione Vita**

<b>Range di Reddito</b>	<b>Promozione Assicurazione vita</b>	<b>Assicurazione Carta di credito</b>	<b>Sesso</b>
40-50K	No	No	M
30-40K	Si	No	F
40-50K	No	No	M
30-40K	Si	Si	M
50-60K	Si	No	F
20-30K	No	No	F
30-40K	Si	Si	M
20-30K	No	No	M
30-40K	No	No	M
30-40K	Si	No	F
40-50K	Si	No	F
20-30K	Si	No	M
50-60K	Si	No	F
40-50K	No	No	M
20-30K	Si	Si	F