Scuola universitaria professionale della Svizzera italiana
Bachelor di Ingegneria Informatica

SUPSI

Machine Learning

Lezione 8 - Clustering

Loris Cannelli, Ricercatore, IDSIA-SUPSI
loris.cannelli@supsi.ch

IDSIA-SUPSI, Polo universitario Lugano - Dipartimento Tecnologie Innovative

Unsupervised Learning

- Immaginiamo di avere N oggetti/dati $\mathbf{x}_1, \ldots, \mathbf{x}_N$, ognuno di dimensione D

Unsupervised Learning

- Immaginiamo di avere N oggetti/dati $\mathbf{x}_1, \ldots, \mathbf{x}_N$, ognuno di dimensione D
- Ipotizziamo che esistano K diverse classi di appartenenza e che ogni oggetto $\mathbf{x}_n$ sia associato ad una delle classi

Unsupervised Learning

- Immaginiamo di avere N oggetti/dati $\mathbf{x}_1, \ldots, \mathbf{x}_N$, ognuno di dimensione D
- Ipotizziamo che esistano K diverse classi di appartenenza e che ogni oggetto $\mathbf{x}_n$ sia associato ad una delle classi
- A differenza di quello che abbiamo visto nelle lezioni passate (*supervised learning*), ora non conosciamo a priori la classe di appartenenza dei dati $\Rightarrow$ siamo noi a volerli classificare! (*unsupervised learning*)

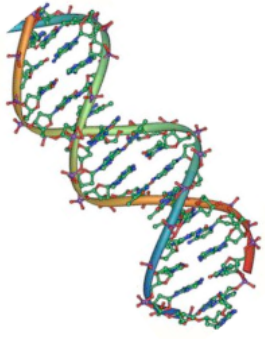
Unsupervised Learning

- Immaginiamo di avere N oggetti/dati $\mathbf{x}_1, \ldots, \mathbf{x}_N$, ognuno di dimensione D
- Ipotizziamo che esistano K diverse classi di appartenenza e che ogni oggetto $\mathbf{x}_n$ sia associato ad una delle classi
- A differenza di quello che abbiamo visto nelle lezioni passate (*supervised learning*), ora non conosciamo a priori la classe di appartenenza dei dati $\Rightarrow$ siamo noi a volerli classificare! (*unsupervised learning*)

$\Rightarrow$ Clustering

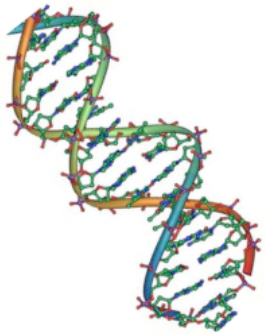
Clustering

- Vogliamo identificare dei pattern o delle similarità



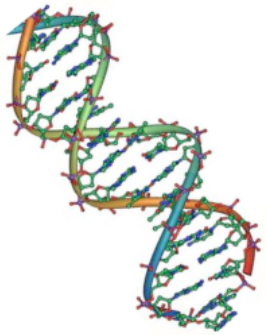
Clustering

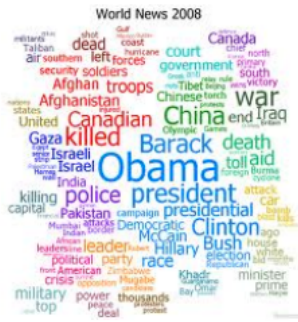
- Vogliamo identificare dei pattern o delle similarità
- I metodi di clustering suddividono osservazioni in gruppi omogenei, chiamati cluster



Clustering

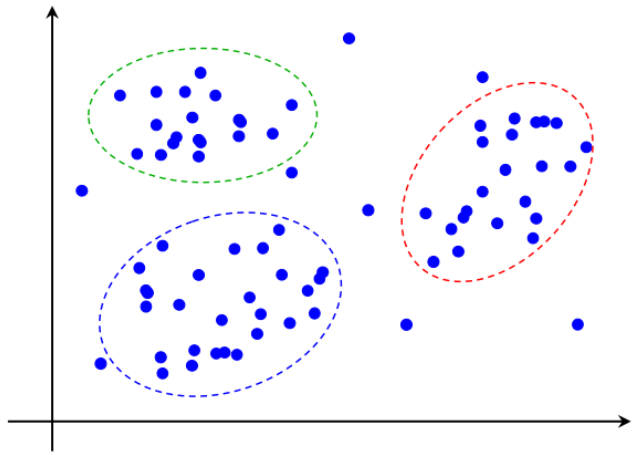
- Vogliamo identificare dei pattern o delle similarità
- I metodi di clustering suddividono osservazioni in gruppi omogenei, chiamati cluster
- I cluster sono costruiti in modo tale che:
  - osservazioni simili appartengono allo stesso cluster
  - osservazioni dissimili appartengono a cluster diversi





Clustering: esempi

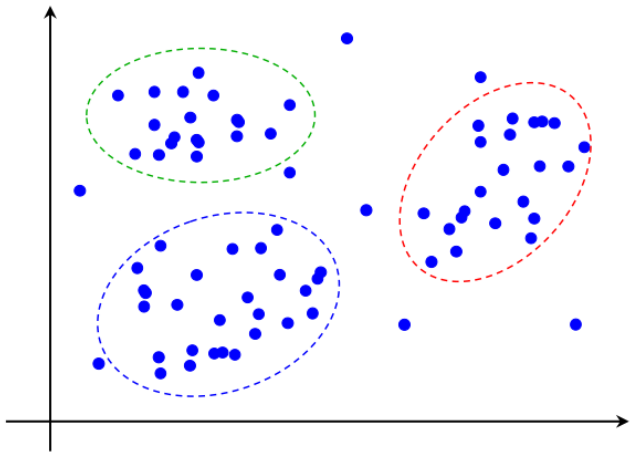
- I metodi di clustering vengono utilizzati per analizzare un fenomeno:
  - Il raggruppamento di clienti in base allo storico dei loro acquisti può permettere di identificare segmenti di mercato per eventuali operazioni di marketing
  - I geni possono essere raggruppati in classi diverse a seconda di come interagiscono con certe molecole. Un punto di partenza per futuri studi è assumere che geni nella stessa classe hanno funzioni simili



Clustering: esempi

- I metodi di clustering vengono utilizzati per analizzare un fenomeno:
  - Il raggruppamento di clienti in base allo storico dei loro acquisti può permettere di identificare segmenti di mercato per eventuali operazioni di marketing
  - I geni possono essere raggruppati in classi diverse a seconda di come interagiscono con certe molecole. Un punto di partenza per futuri studi è assumere che geni nella stessa classe hanno funzioni simili

- I cluster possono anche essere utilizzati per preparare i dati a successive fasi di data mining: la separazione in cluster può permettere lo sviluppo di modelli di classificazione specifici per ogni cluster

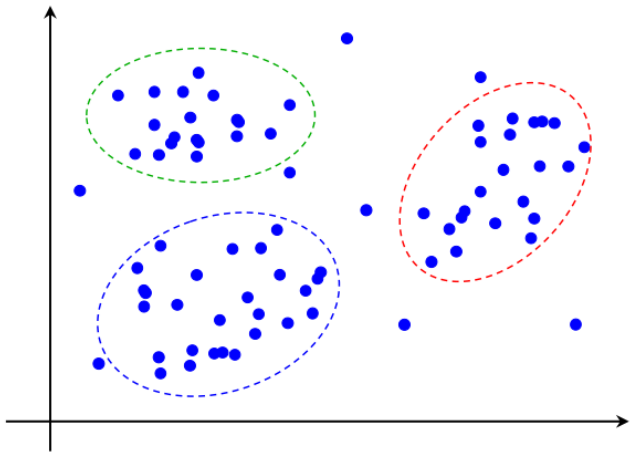


Clustering: esempi

- I metodi di clustering vengono utilizzati per analizzare un fenomeno:
  - Il raggruppamento di clienti in base allo storico dei loro acquisti può permettere di identificare segmenti di mercato per eventuali operazioni di marketing
  - I geni possono essere raggruppati in classi diverse a seconda di come interagiscono con certe molecole. Un punto di partenza per futuri studi è assumere che geni nella stessa classe hanno funzioni simili

- I cluster possono anche essere utilizzati per preparare i dati a successive fasi di data mining: la separazione in cluster può permettere lo sviluppo di modelli di classificazione specifici per ogni cluster

- I cluster possono aiutare a evidenziare osservazioni anomale (outlier) nella fase di pulizia del dataset



Clustering: complessità

- Il problema ha complessità NP hard: la soluzione ottima può essere trovata con sicurezza solo tramite ricerca esaustiva

Clustering: complessità

- Il problema ha complessità NP hard: la soluzione ottima può essere trovata con sicurezza solo tramite ricerca esaustiva
- Il numero di cluster possibili è dell'ordine di $\frac{K^N}{K!}$ (numero di Stirling di seconda specie)
  Esempio: $N = 100$, $K = 5$, possibili cluster $\sim 10^{67}$

Clustering: complessità

- Il problema ha complessità NP hard: la soluzione ottima può essere trovata con sicurezza solo tramite ricerca esaustiva
- Il numero di cluster possibili è dell'ordine di $\frac{K^N}{K!}$ (numero di Stirling di seconda specie)
  Esempio: $N = 100$, $K = 5$, possibili cluster $\sim 10^{67}$
- Se K non è noto a priori, il numero di possibili cluster è dato dalla somma per K che va da 1 a N dei numeri precedenti (numero di Bell)

Clustering: complessità

- Il problema ha complessità NP hard: la soluzione ottima può essere trovata con sicurezza solo tramite ricerca esaustiva
- Il numero di cluster possibili è dell'ordine di $\frac{K^N}{K!}$ (numero di Stirling di seconda specie)
  Esempio: $N = 100$, $K = 5$, possibili cluster $\sim 10^{67}$
- Se K non è noto a priori, il numero di possibili cluster è dato dalla somma per K che va da 1 a N dei numeri precedenti (numero di Bell)
- Nel caso $K = 2$ si dimostra facilmente che i possibili cluster sono $2^{K-1} - 1$
  Esempio: $X = \{A, B, C, D\}$; $K = 2$; Cluster:
  $(A)(B, C, D), (B)(A, C, D), (C)(A, B, D), (D)(A, B, C), (A, B)(C, D),$
  $(A, C)(B, D), (A, D)(B, C)$

Clustering: criteri

- Quale criterio seguire?

Clustering: criteri

- Quale criterio seguire?

1. i pattern all'interno dello stesso cluster devono essere tra loro più simili rispetto a pattern appartenenti a cluster diversi

Clustering: criteri

- Quale criterio seguire?

1. i pattern all'interno dello stesso cluster devono essere tra loro più simili rispetto a pattern appartenenti a cluster diversi
2. i cluster sono costituiti da nuvole di punti a densità relativamente elevata separate da zone dove la densità e più bassa

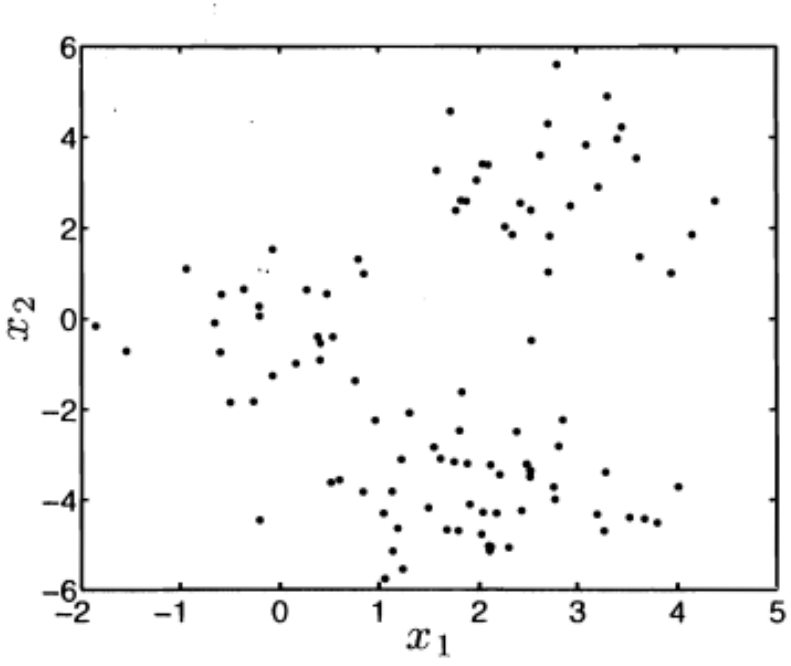
Clustering: criteri

- Quale criterio seguire?

1. i pattern all'interno dello stesso cluster devono essere tra loro più simili rispetto a pattern appartenenti a cluster diversi
2. i cluster sono costituiti da nuvole di punti a densità relativamente elevata separate da zone dove la densità e più bassa

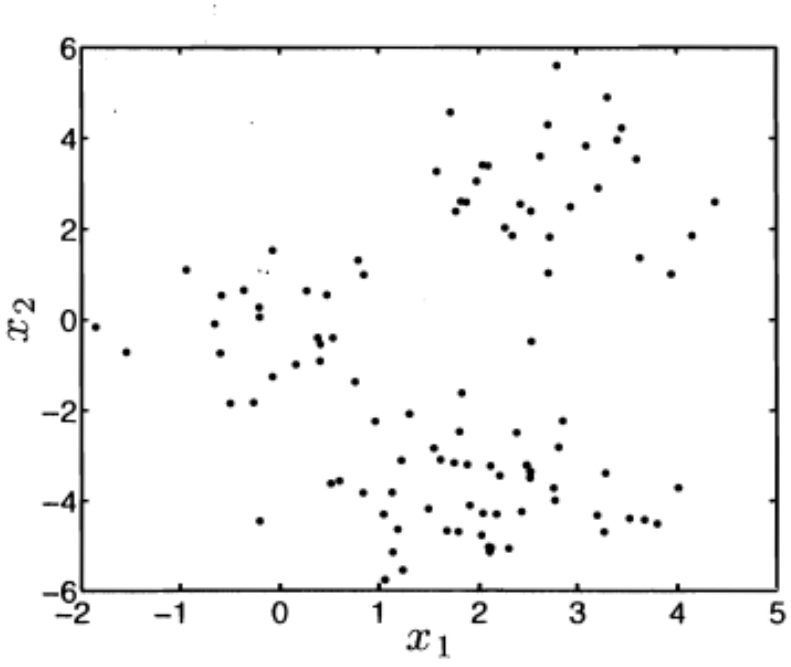
- E' necessaria una qualche definizione di distanza

Clustering: criteri



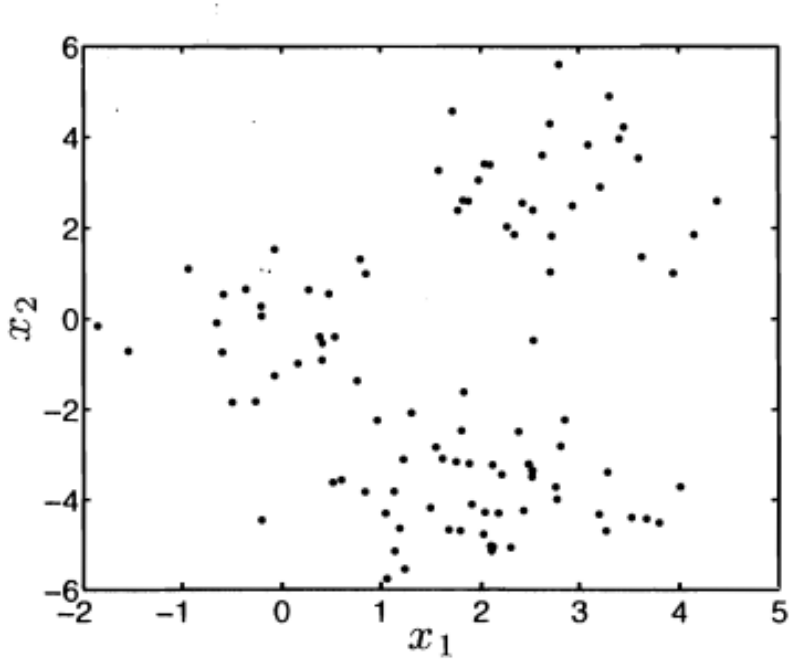
- Riusciamo chiaramente a vedere nella figura tre cluster principali

Clustering: criteri



- Riusciamo chiaramente a vedere nella figura tre cluster principali
- Quello che stiamo facendo è applicare la nozione di distanza Euclidea

Clustering: criteri



- Riusciamo chiaramente a vedere nella figura tre cluster principali
- Quello che stiamo facendo è applicare la nozione di distanza Euclidea
- $\mathbf{x}_i$ è simile a $\mathbf{x}_j$ se $(\mathbf{x}_{i1} - \mathbf{x}_{j1})^2 + (\mathbf{x}_{i2} - \mathbf{x}_{j2})^2 = (\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j)$ è basso

Clustering: distanze

Esistono diversi tipi di distanze e alcune sono più adatte di altre per specifiche applicazioni pratiche

Clustering: distanze

Esistono diversi tipi di distanze e alcune sono più adatte di altre per specifiche applicazioni pratiche

- Distanza Minkowski: $d_q \triangleq \sqrt[q]{\sum_{t=1}^{D} |\mathbf{x}_{it} - \mathbf{x}_{jt}|^q} \rightarrow$ per $q = 2$ otteniamo la Distanza Euclidea

Clustering: distanze

Esistono diversi tipi di distanze e alcune sono più adatte di altre per specifiche applicazioni pratiche

- Distanza Minkowski: $d_q \triangleq \sqrt[q]{\sum_{t=1}^{D} |\mathbf{x}_{it} - \mathbf{x}_{jt}|^q} \rightarrow$ per $q = 2$ otteniamo la Distanza Euclidea
- Distanza Manhattan ($q = 1$): $d_M \triangleq \sum_{t=1}^{D} |\mathbf{x}_{it} - \mathbf{x}_{jt}|$

Clustering: distanze

Esistono diversi tipi di distanze e alcune sono più adatte di altre per specifiche applicazioni pratiche

- Distanza Minkowski: $d_q \triangleq \sqrt[q]{\sum_{t=1}^{D} |\mathbf{x}_{it} - \mathbf{x}_{jt}|^q} \rightarrow$ per $q = 2$ otteniamo la Distanza Euclidea
- Distanza Manhattan ($q = 1$): $d_M \triangleq \sum_{t=1}^{D} |\mathbf{x}_{it} - \mathbf{x}_{jt}|$
- Distanza Chebyshev ($q = +\infty$): $d_C \triangleq \max_{t=1,\ldots,D} |\mathbf{x}_{it} - \mathbf{x}_{jt}|$

Clustering: distanze

Esistono diversi tipi di distanze e alcune sono più adatte di altre per specifiche applicazioni pratiche

- Distanza Minkowski: $d_q \triangleq \sqrt[q]{\sum_{t=1}^{D} |\mathbf{x}_{it} - \mathbf{x}_{jt}|^q} \rightarrow$ per $q = 2$ otteniamo la Distanza Euclidea
- Distanza Manhattan ($q = 1$): $d_M \triangleq \sum_{t=1}^{D} |\mathbf{x}_{it} - \mathbf{x}_{jt}|$
- Distanza Chebyshev ($q = +\infty$): $d_C \triangleq \max_{t=1,\ldots,D} |\mathbf{x}_{it} - \mathbf{x}_{jt}|$
- Distanza normalizzata/standardizzata: $d_w \triangleq \sqrt[q]{\sum_{t=1}^{D} w_t |\mathbf{x}_{it} - \mathbf{x}_{jt}|^q}$, dove w_t è un peso inversamente proporzionale al valore massimo che l'attributo assume

Clustering: distanze

Esistono diversi tipi di distanze e alcune sono più adatte di altre per specifiche applicazioni pratiche

- Distanza Minkowski: $d_q \triangleq \sqrt[q]{\sum_{t=1}^{D} |\mathbf{x}_{it} - \mathbf{x}_{jt}|^q} \rightarrow$ per $q = 2$ otteniamo la Distanza Euclidea
- Distanza Manhattan ($q = 1$): $d_M \triangleq \sum_{t=1}^{D} |\mathbf{x}_{it} - \mathbf{x}_{jt}|$
- Distanza Chebyshev ($q = +\infty$): $d_C \triangleq \max_{t=1,\ldots,D} |\mathbf{x}_{it} - \mathbf{x}_{jt}|$
- Distanza normalizzata/standardizzata: $d_w \triangleq \sqrt[q]{\sum_{t=1}^{D} w_t |\mathbf{x}_{it} - \mathbf{x}_{jt}|^q}$, dove w_t è un peso inversamente proporzionale al valore massimo che l'attributo assume
- Esistono distanze anche per valori nominali. La più semplice consiste nell'associare 1 a attributi identici e 0 ad attributi diversi

Clustering: centroidi

- Ipotizziamo che ogni cluster abbia un proprio *centroide* $\boldsymbol{\mu}_k, k = 1, \ldots, K$

Clustering: centroidi

- Ipotizziamo che ogni cluster abbia un proprio *centroide* $\boldsymbol{\mu}_k, k = 1, \ldots, K$
- Usiamo $z_{nk} = 1$ per indicare che l'oggetto n appartiene al cluster k e $z_{nk} = 0$ per indicare che non appartiene a quel cluster

Clustering: centroidi

- Ipotizziamo che ogni cluster abbia un proprio *centroide* $\boldsymbol{\mu}_k, k = 1, \ldots, K$
- Usiamo $z_{nk} = 1$ per indicare che l'oggetto n appartiene al cluster k e $z_{nk} = 0$ per indicare che non appartiene a quel cluster
- Immaginiamo di avere già classificato secondo qualche criterio gli N oggetti a nostra disposizione in K cluster. La formula del centroide si ottiene quindi come:

$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} z_{nk} \mathbf{x}_n}{\sum_{n=1}^{N} z_{nk}}$$

Clustering: centroidi

- Ipotizziamo che ogni cluster abbia un proprio *centroide* $\boldsymbol{\mu}_k, k = 1, \ldots, K$
- Usiamo $z_{nk} = 1$ per indicare che l'oggetto n appartiene al cluster k e $z_{nk} = 0$ per indicare che non appartiene a quel cluster
- Immaginiamo di avere già classificato secondo qualche criterio gli N oggetti a nostra disposizione in K cluster. La formula del centroide si ottiene quindi come:

$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} z_{nk}\mathbf{x}_n}{\sum_{n=1}^{N} z_{nk}}$$

- Siamo bloccati in un loop: vogliamo assegnare ogni oggetto al centroide più vicino, ma per calcolare la posizione di un centroide dobbiamo prima sapere quali oggetti sono assegnati ad esso

$K-$means

L'algoritmo $K-$means risolve proprio il problema di cui stiamo parlando:

$K-$means

L'algoritmo $K-$means risolve proprio il problema di cui stiamo parlando:

1. Scegli il valore di K, quale distanza utilizzare e il numero di iterazioni massime da eseguire

$K-$means

L'algoritmo $K-$means risolve proprio il problema di cui stiamo parlando:

1. Scegli il valore di K, quale distanza utilizzare e il numero di iterazioni massime da eseguire
2. Inizializza i valori iniziali dei centroidi $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K$

$K-$means

L'algoritmo $K-$means risolve proprio il problema di cui stiamo parlando:

1. Scegli il valore di K, quale distanza utilizzare e il numero di iterazioni massime da eseguire
2. Inizializza i valori iniziali dei centroidi $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K$
3. Assegna ogni oggetto $\mathbf{x}_1, \ldots, \mathbf{x}_N$ al centroide più vicino

$K-$means

L'algoritmo $K-$means risolve proprio il problema di cui stiamo parlando:

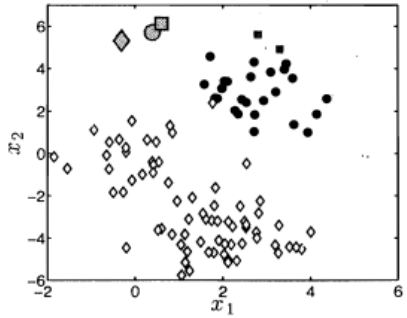
1. Scegli il valore di K, quale distanza utilizzare e il numero di iterazioni massime da eseguire
2. Inizializza i valori iniziali dei centroidi $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K$
3. Assegna ogni oggetto $\mathbf{x}_1, \ldots, \mathbf{x}_N$ al centroide più vicino
4. Calcola i valori dei nuovi centroidi

$K-$means

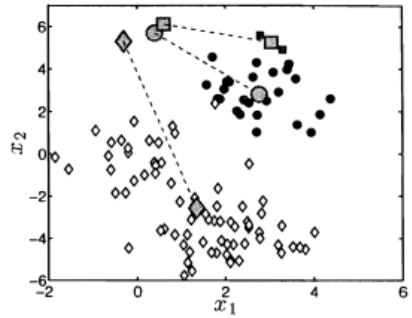
L'algoritmo $K-$means risolve proprio il problema di cui stiamo parlando:

1. Scegli il valore di K, quale distanza utilizzare e il numero di iterazioni massime da eseguire
2. Inizializza i valori iniziali dei centroidi $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K$
3. Assegna ogni oggetto $\mathbf{x}_1, \ldots, \mathbf{x}_N$ al centroide più vicino
4. Calcola i valori dei nuovi centroidi
5. Riparti dal punto 3, a meno che
   - Le assegnazioni dei dati sono rimaste invariate in due iterazioni successive
   - Il numero di iterazioni massimo è stato raggiunto

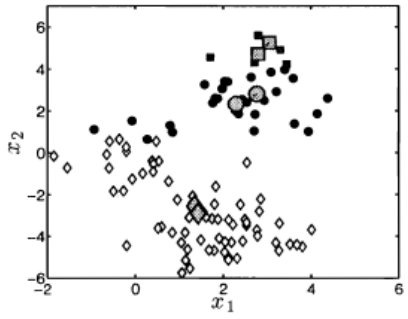
$K-$means: esempio



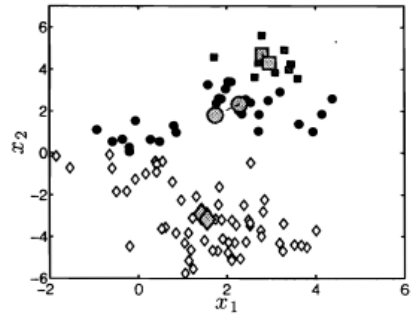
(a) Data and initial random means. Means are depicted by large symbols. Each data object is given the symbol of its closest mean



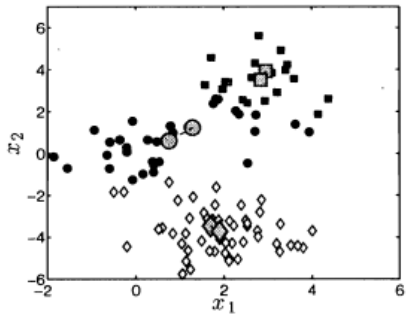
(b) Means updated according to assigned objects



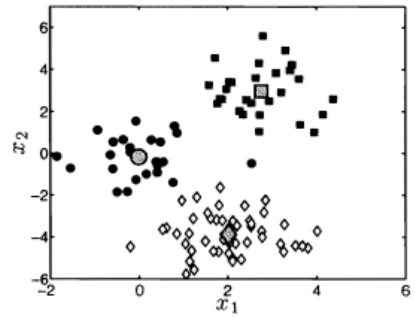
(c) Objects re-assigned to new means and means updated again



(d) Means updated after three iterations



(e) Means updated after five iterations



(f) Means updated after eight iterations. Algorithm has converged

$K-$means: analisi

- Si può dimostrare che l'algoritmo $K-$means converge a un minimo *locale* della funzione costo:

$$L \triangleq \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} (\mathbf{x}_n - \boldsymbol{\mu}_k)^T (\mathbf{x}_n - \boldsymbol{\mu}_k)$$

  il cui valore equivale alla somma delle distanze di ogni oggetto dal centroide di appartenenza

$K-$means: analisi

- Si può dimostrare che l'algoritmo $K-$means converge a un minimo *locale* della funzione costo:

$$L \triangleq \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} (\mathbf{x}_n - \boldsymbol{\mu}_k)^T (\mathbf{x}_n - \boldsymbol{\mu}_k)$$

  il cui valore equivale alla somma delle distanze di ogni oggetto dal centroide di appartenenza

- L'algoritmo converge a soluzioni diverse a seconda dell'inizializzazione dei centroidi iniziali

$K-$means: analisi

- Si può dimostrare che l'algoritmo $K-$means converge a un minimo *locale* della funzione costo:

$$L \triangleq \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} (\mathbf{x}_n - \boldsymbol{\mu}_k)^T (\mathbf{x}_n - \boldsymbol{\mu}_k)$$

  il cui valore equivale alla somma delle distanze di ogni oggetto dal centroide di appartenenza

- L'algoritmo converge a soluzioni diverse a seconda dell'inizializzazione dei centroidi iniziali

- Spesso si fa girare l'algoritmo più volte, testando diverse inizializzazioni random e poi si sceglie la soluzione che ha minimizzato di più L

$K-$means: analisi

- Si può dimostrare che l'algoritmo $K-$means converge a un minimo *locale* della funzione costo:

$$L \triangleq \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} (\mathbf{x}_n - \boldsymbol{\mu}_k)^T (\mathbf{x}_n - \boldsymbol{\mu}_k)$$

  il cui valore equivale alla somma delle distanze di ogni oggetto dal centroide di appartenenza

- L'algoritmo converge a soluzioni diverse a seconda dell'inizializzazione dei centroidi iniziali

- Spesso si fa girare l'algoritmo più volte, testando diverse inizializzazioni random e poi si sceglie la soluzione che ha minimizzato di più L

- Un'altra idea è inizializzare i centroidi scegliendo K elementi random del dataset

$K-$means: analisi

- Si può dimostrare che l'algoritmo $K-$means converge a un minimo *locale* della funzione costo:

$$L \triangleq \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} (\mathbf{x}_n - \boldsymbol{\mu}_k)^T (\mathbf{x}_n - \boldsymbol{\mu}_k)$$

  il cui valore equivale alla somma delle distanze di ogni oggetto dal centroide di appartenenza

- L'algoritmo converge a soluzioni diverse a seconda dell'inizializzazione dei centroidi iniziali

- Spesso si fa girare l'algoritmo più volte, testando diverse inizializzazioni random e poi si sceglie la soluzione che ha minimizzato di più L

- Un'altra idea è inizializzare i centroidi scegliendo K elementi random del dataset

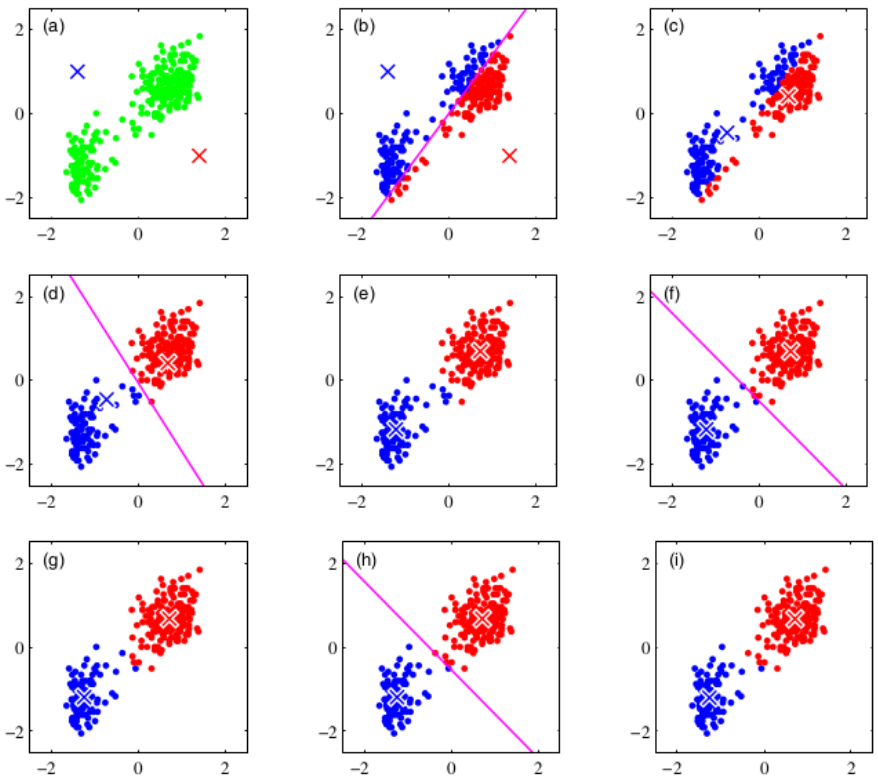
- Spesso si usa la soluzione generata da $K-$means dopo poche iterazioni come inizializzazione per algoritmi di Machine Learning più raffinati

$K-$means: Expectation-Minimization (EM)

- E'stato dimostrato che $K-$means è un'istanza della vasta classe di algoritmi noti come EM

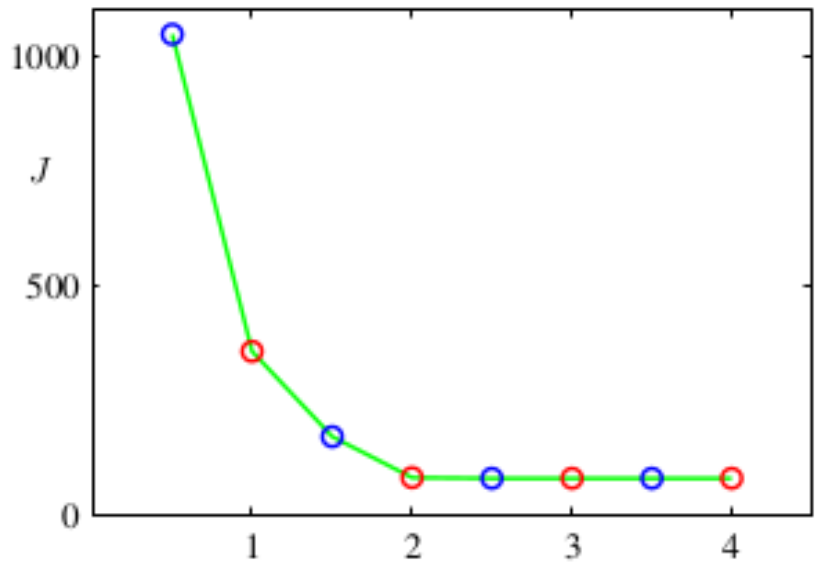
$K-$means: Expectation-Minimization (EM)

- E'stato dimostrato che $K-$means è un'istanza della vasta classe di algoritmi noti come EM
- Step E: assegnare gli oggetti ai centroidi più vicini; Step M: calcolare le nuove posizioni dei centroidi



$K-$means: Expectation-Minimization (EM)

- E'stato dimostrato che $K-$means è un'istanza della vasta classe di algoritmi noti come EM
- Step E: assegnare gli oggetti ai centroidi più vicini; Step M: calcolare le nuove posizioni dei centroidi

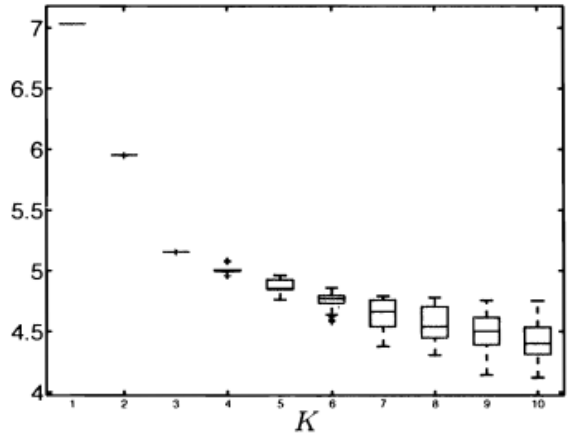


$K-$means: scegliere K

- Il corretto valore di K è fortemente dipendente dal problema che si vuole risolvere

$K-$means: scegliere K

- Il corretto valore di K è fortemente dipendente dal problema che si vuole risolvere

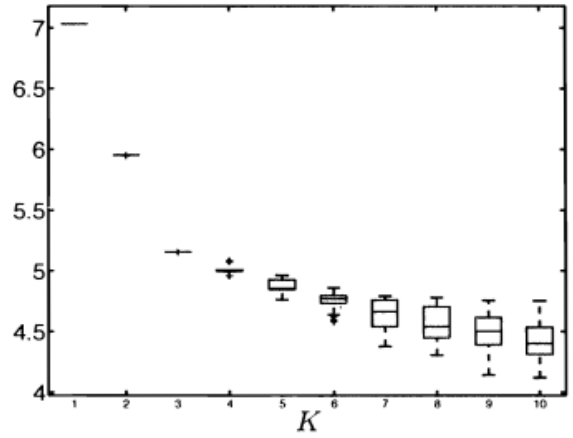


La figura mostra il valore di $\log L$ al variare di K. Per ogni K sono state testate 50 differenti inizializzazioni random

- Aumentare il valore di K fa sempre diminuire L, ma non è detto che si ottenga il clustering più adatto al problema che si vuole risolvere

$K-$means: scegliere K

- Il corretto valore di K è fortemente dipendente dal problema che si vuole risolvere

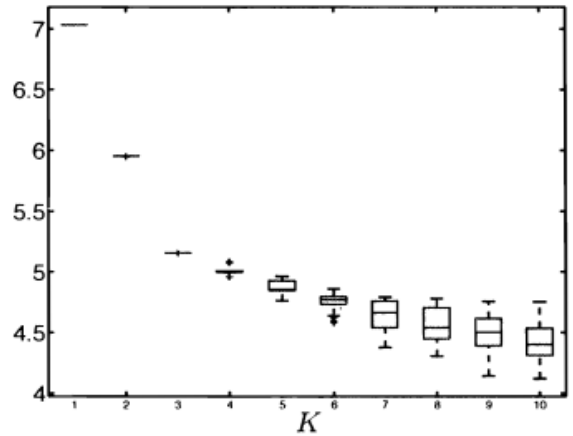


La figura mostra il valore di $\log L$ al variare di K. Per ogni K sono state testate 50 differenti inizializzazioni random

- Aumentare il valore di K fa sempre diminuire L, ma non è detto che si ottenga il clustering più adatto al problema che si vuole risolvere
- Ad esempio, nel caso limite in cui si sceglie $K = N$ si ottiene $L = 0$, ma il clustering è privo di senso

$K-$means: scegliere K

- Il corretto valore di K è fortemente dipendente dal problema che si vuole risolvere



La figura mostra il valore di $\log L$ al variare di K. Per ogni K sono state testate 50 differenti inizializzazioni random

- Aumentare il valore di K fa sempre diminuire L, ma non è detto che si ottenga il clustering più adatto al problema che si vuole risolvere
- Ad esempio, nel caso limite in cui si sceglie $K = N$ si ottiene $L = 0$, ma il clustering è privo di senso
- Un buon approccio è scegliere il valore di K che produce una forte discontinuità/scalino nel grafico di L

$K-$means: scegliere K

$K-$means: altre considerazioni

- $K-$means è un algoritmo semplice, ma produce buone soluzioni

$K-$means: altre considerazioni

- $K-$means è un algoritmo semplice, ma produce buone soluzioni
- Il valore di K e l'inizializzazione dei centroidi sono parametri cruciali

$K-$means: altre considerazioni

- $K-$means è un algoritmo semplice, ma produce buone soluzioni
- Il valore di K e l'inizializzazione dei centroidi sono parametri cruciali
- Solitamente la convergenza si ottiene in poche iterazioni (< 10)

$K-$means: altre considerazioni

- $K-$means è un algoritmo semplice, ma produce buone soluzioni
- Il valore di K e l'inizializzazione dei centroidi sono parametri cruciali
- Solitamente la convergenza si ottiene in poche iterazioni (< 10)
- Utilizzando la distanza Euclidea si ottengono cluster radiali/sferici. Con altre distanze si ottengono geometrie diverse

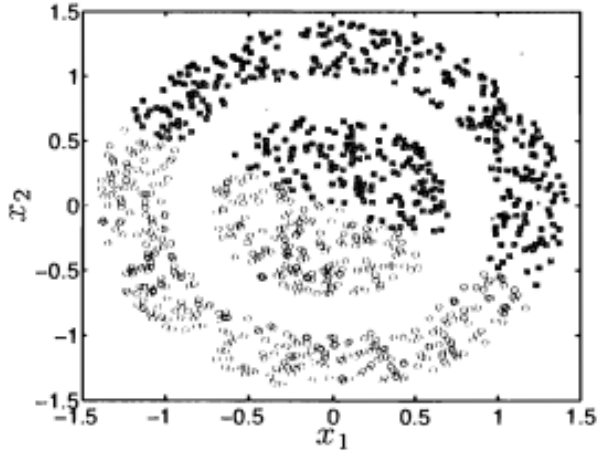
$K-$means: altre considerazioni

- $K-$means è un algoritmo semplice, ma produce buone soluzioni
- Il valore di K e l'inizializzazione dei centroidi sono parametri cruciali
- Solitamente la convergenza si ottiene in poche iterazioni (< 10)
- Utilizzando la distanza Euclidea si ottengono cluster radiali/sferici. Con altre distanze si ottengono geometrie diverse
- Esiste una versione *online* dell'algoritmo: si riceve un dato alla volta, lo si associa al centroide più vicino e si aggiorna il centroide secondo la regola:

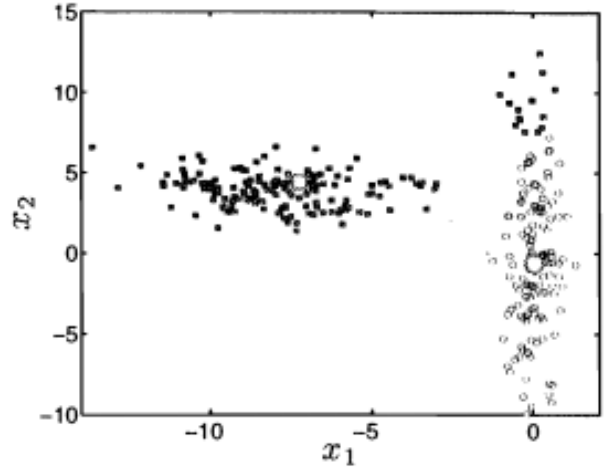
$$\boldsymbol{\mu}_k^{\text{new}} = \boldsymbol{\mu}_k^{\text{old}} + \eta_n \left(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{old}} \right),$$

dove $\eta_n \in (0, 1]$ è la *learning rate*, che solitamente è monotonicamente decrescente

$K-$means: svantaggi

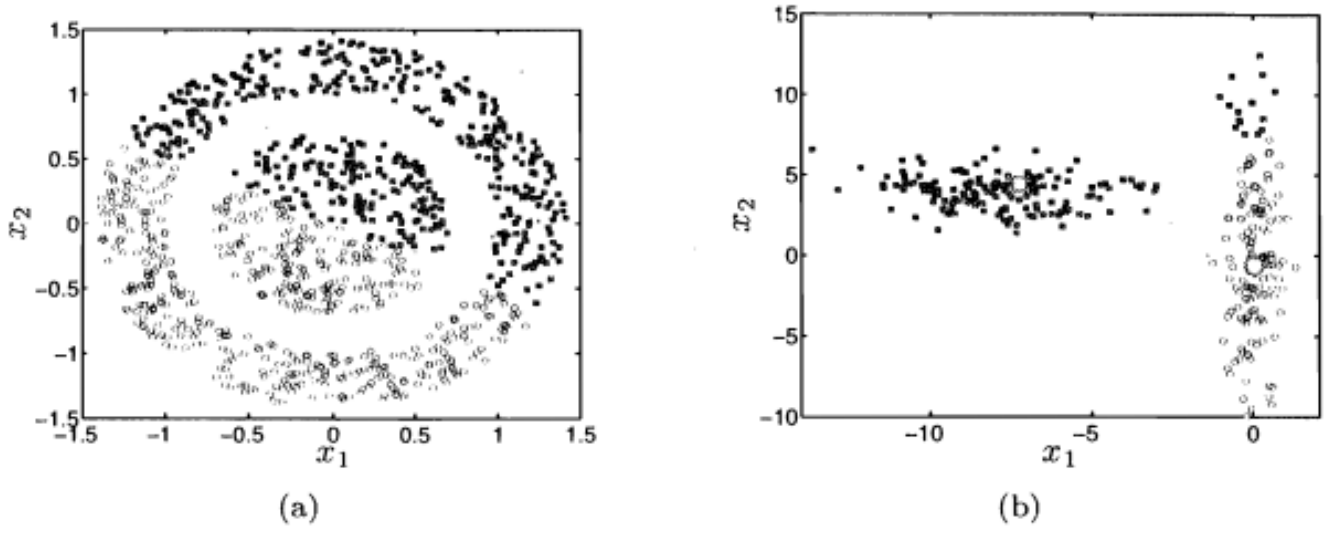


(a)



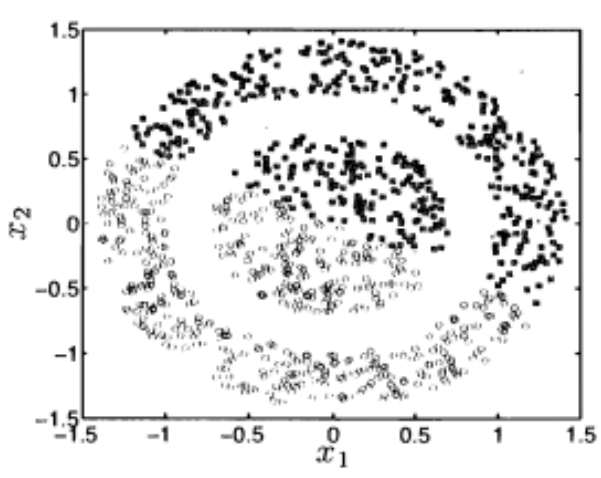
(b)

$K-$means: svantaggi

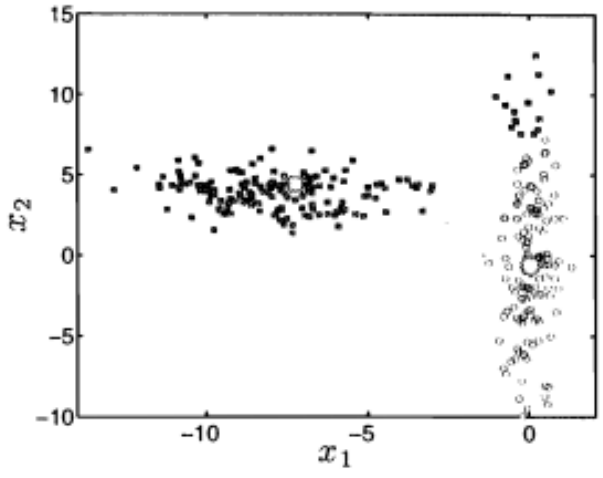


Difficoltà ad apprendere geometrie particolari

$K-$means: svantaggi



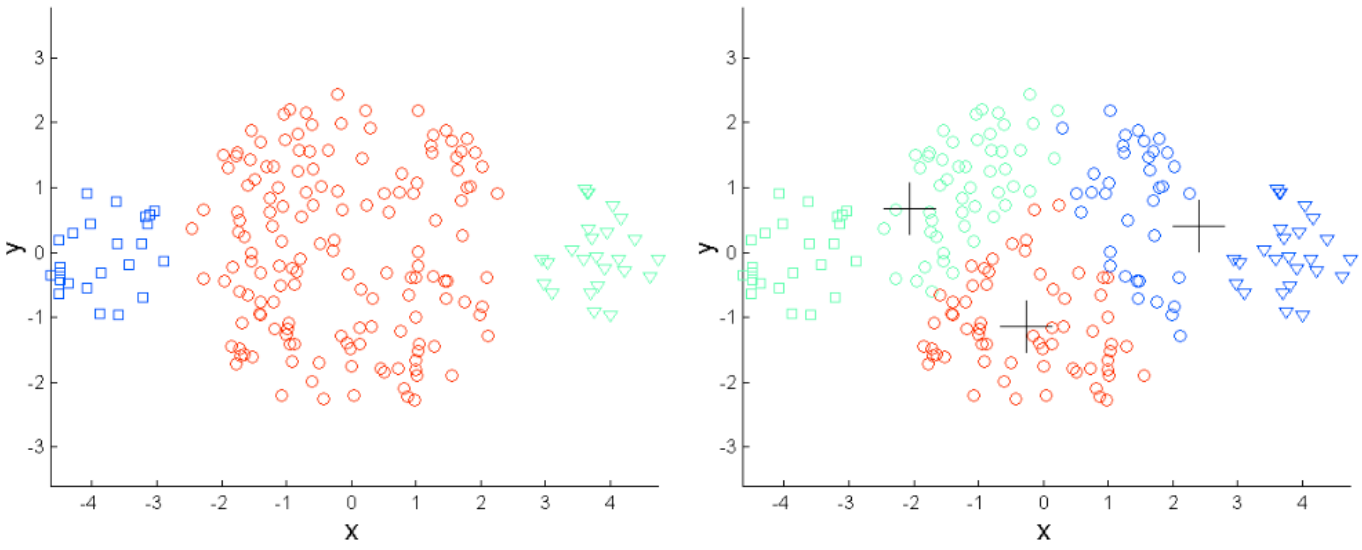
(a)



(b)

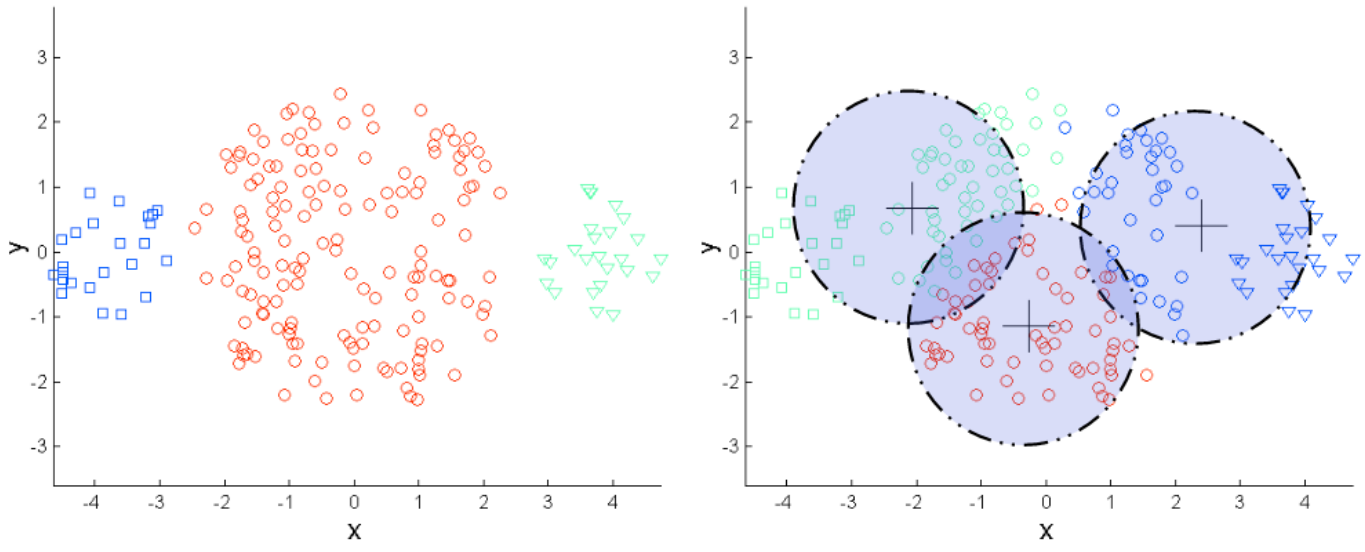
possibile soluzione: aumentare K $\Rightarrow$ si identificano parti dei cluster che devono successivamente essere aggregate.

$K-$means: svantaggi



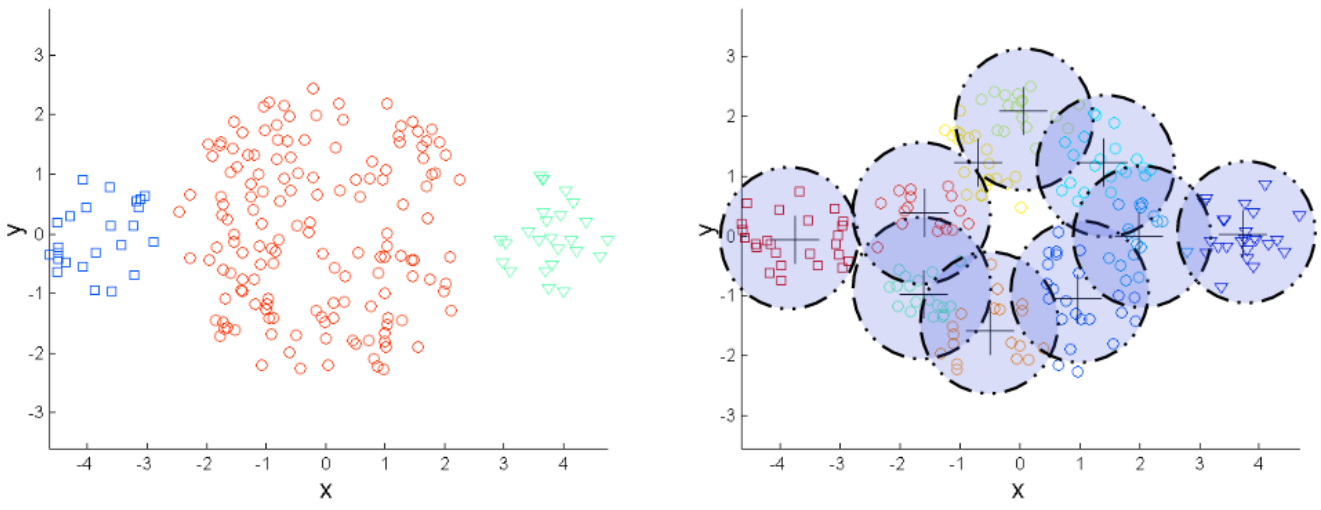
possibile soluzione: aumentare K $\Rightarrow$ si identificano parti dei cluster che devono successivamente essere aggregate.

$K-$means: svantaggi



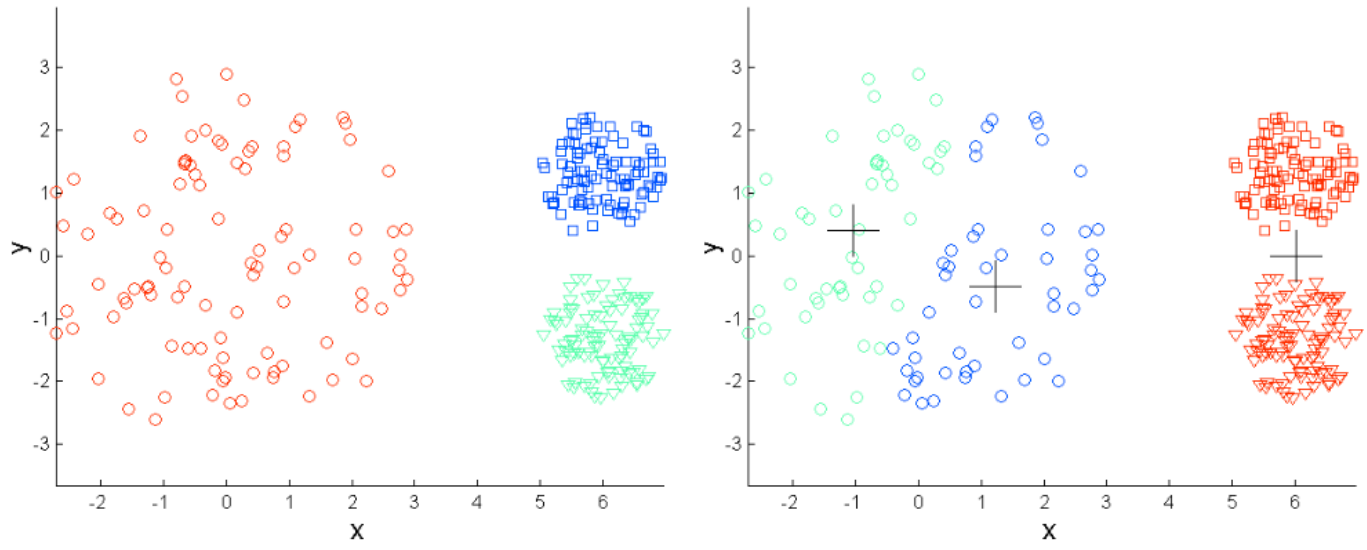
possibile soluzione: aumentare K $\Rightarrow$ si identificano parti dei cluster che devono successivamente essere aggregate.

$K-$means: svantaggi



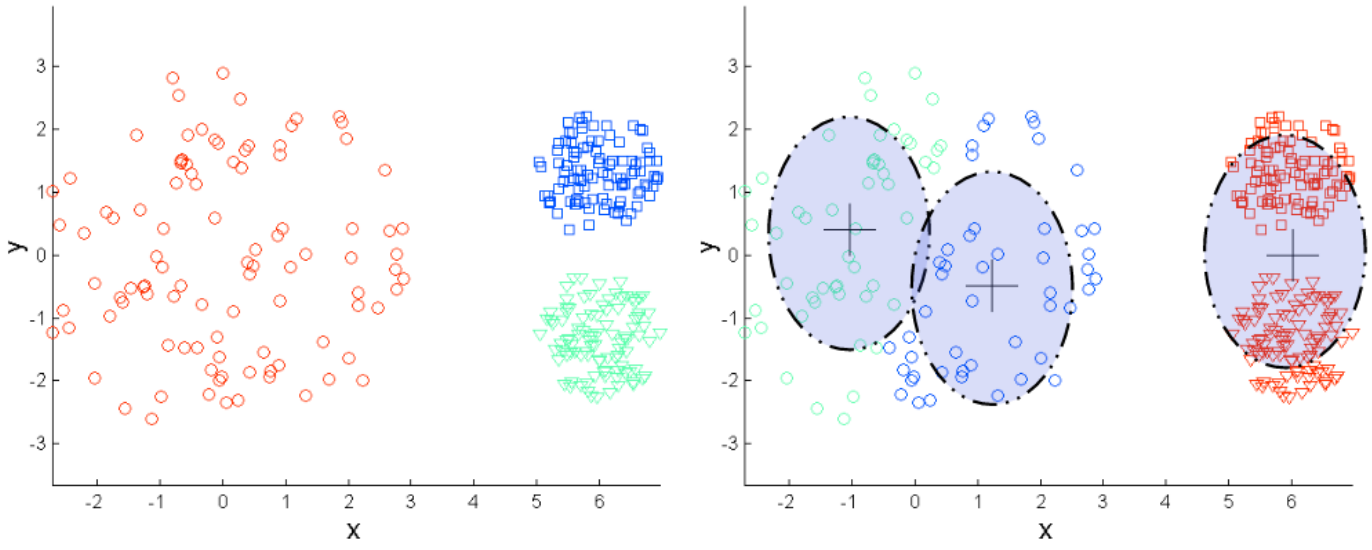
possibile soluzione: aumentare K $\Rightarrow$ si identificano parti dei cluster che devono successivamente essere aggregate.

$K-$means: svantaggi



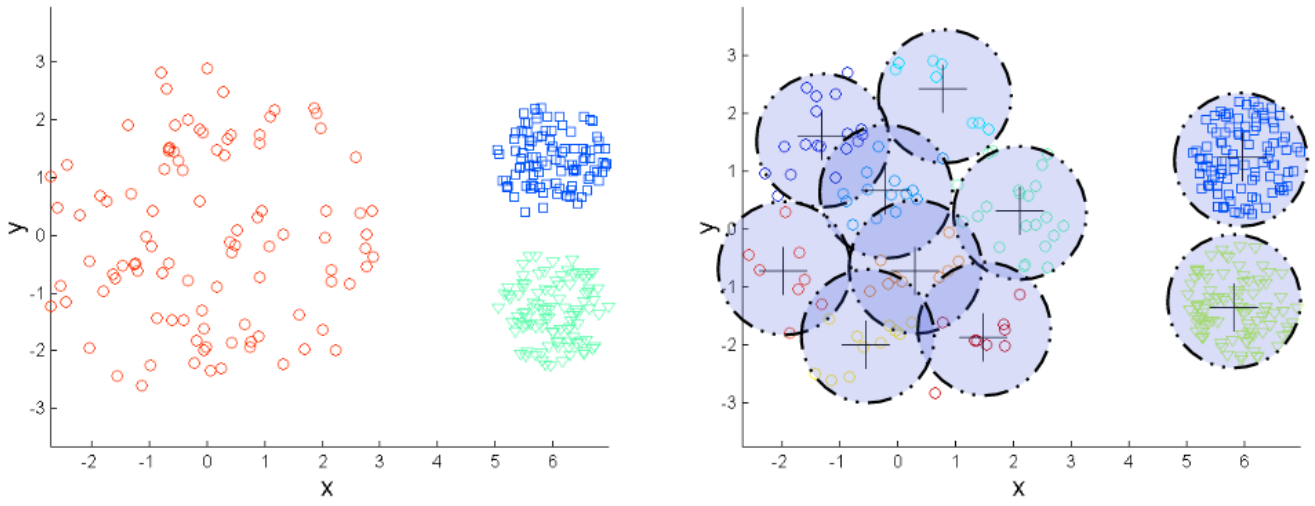
possibile soluzione: aumentare $K \Rightarrow$ si identificano parti dei cluster che devono successivamente essere aggregate.

$K-$means: svantaggi



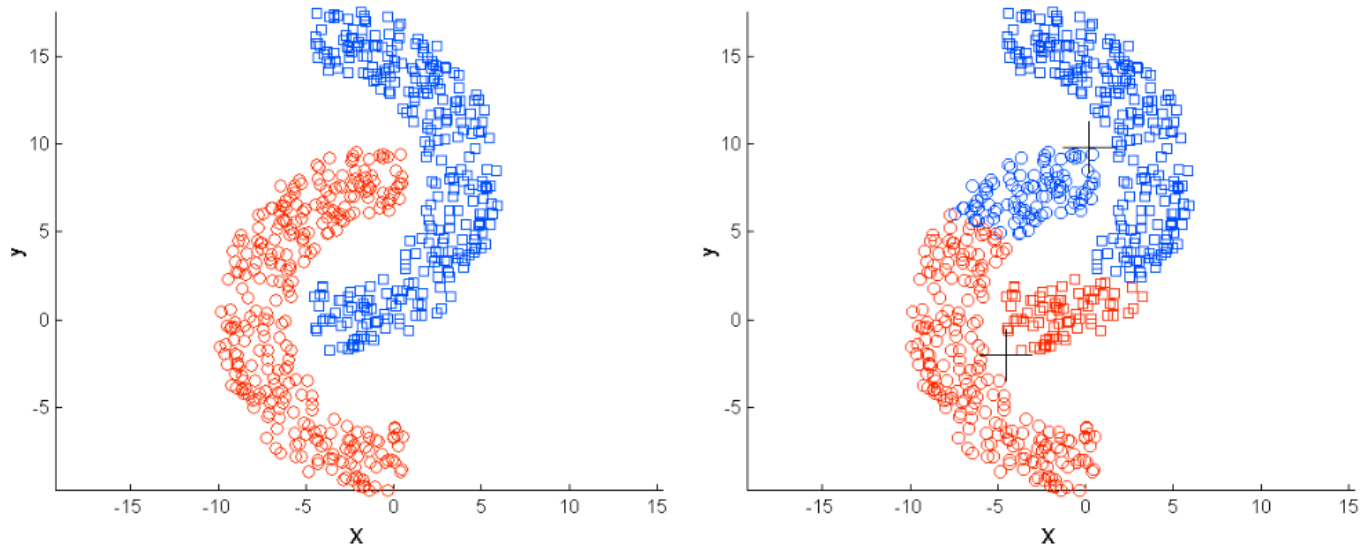
possibile soluzione: aumentare K $\Rightarrow$ si identificano parti dei cluster che devono successivamente essere aggregate.

$K-$means: svantaggi



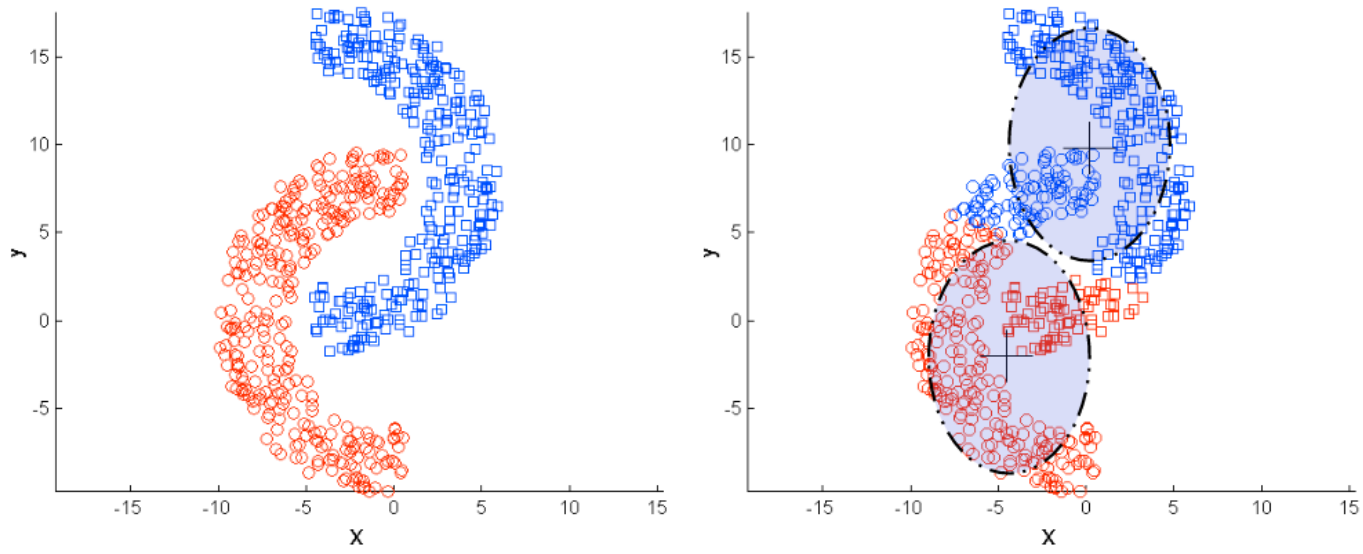
possibile soluzione: aumentare K $\Rightarrow$ si identificano parti dei cluster che devono successivamente essere aggregate.

$K-$means: svantaggi



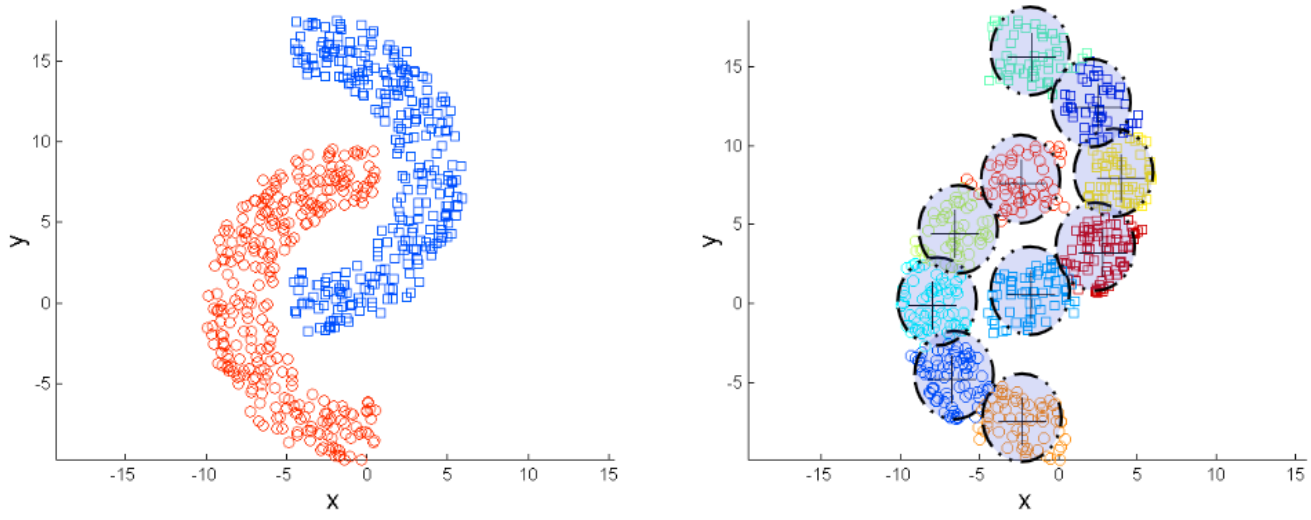
possibile soluzione: aumentare $K \Rightarrow$ si identificano parti dei cluster che devono successivamente essere aggregate.

$K-$means: svantaggi



possibile soluzione: aumentare $K \Rightarrow$ si identificano parti dei cluster che devono successivamente essere aggregate.

$K-$means: svantaggi



possibile soluzione: aumentare K $\Rightarrow$ si identificano parti dei cluster che devono successivamente essere aggregate.

Categorie di clustering

Categorie di clustering

- Metodi di partizione: suddividono il dataset in un numero definito di cluster non vuoti sulla base delle distanze tra le osservazioni ($\rightarrow K-$means)

Categorie di clustering

- Metodi di partizione: suddividono il dataset in un numero definito di cluster non vuoti sulla base delle distanze tra le osservazioni ($\rightarrow K-$means)
- Metodi gerarchici: ricavano molteplici suddivisioni in cluster, sfruttando la struttura ad albero e utilizzando valori di soglia differenti

Categorie di clustering

- Metodi di partizione: suddividono il dataset in un numero definito di cluster non vuoti sulla base delle distanze tra le osservazioni ($\rightarrow K-$means)
- Metodi gerarchici: ricavano molteplici suddivisioni in cluster, sfruttando la struttura ad albero e utilizzando valori di soglia differenti
- Metodi basati sulla densità: sfruttano la densità locale per un intorno delle osservazioni, ragionano su un diametro specificato in modo tale che esso contenga un numero di osservazioni non inferiore ad una certa soglia fornita dal decision maker, identificano cluster con forma non convessa e sono in grado di isolare gli outlier

Categorie di clustering

- Metodi di partizione: suddividono il dataset in un numero definito di cluster non vuoti sulla base delle distanze tra le osservazioni ($\rightarrow K-$means)
- Metodi gerarchici: ricavano molteplici suddivisioni in cluster, sfruttando la struttura ad albero e utilizzando valori di soglia differenti
- Metodi basati sulla densità: sfruttano la densità locale per un intorno delle osservazioni, ragionano su un diametro specificato in modo tale che esso contenga un numero di osservazioni non inferiore ad una certa soglia fornita dal decision maker, identificano cluster con forma non convessa e sono in grado di isolare gli outlier

Attribuzioni

Categorie di clustering

- Metodi di partizione: suddividono il dataset in un numero definito di cluster non vuoti sulla base delle distanze tra le osservazioni ($\rightarrow K-$means)
- Metodi gerarchici: ricavano molteplici suddivisioni in cluster, sfruttando la struttura ad albero e utilizzando valori di soglia differenti
- Metodi basati sulla densità: sfruttano la densità locale per un intorno delle osservazioni, ragionano su un diametro specificato in modo tale che esso contenga un numero di osservazioni non inferiore ad una certa soglia fornita dal decision maker, identificano cluster con forma non convessa e sono in grado di isolare gli outlier

Attribuzioni

- Attribuzione esclusiva: ogni osservazione è assegnata ad un solo cluster ($\rightarrow K-$means)

Categorie di clustering

- Metodi di partizione: suddividono il dataset in un numero definito di cluster non vuoti sulla base delle distanze tra le osservazioni ($\rightarrow K-$means)
- Metodi gerarchici: ricavano molteplici suddivisioni in cluster, sfruttando la struttura ad albero e utilizzando valori di soglia differenti
- Metodi basati sulla densità: sfruttano la densità locale per un intorno delle osservazioni, ragionano su un diametro specificato in modo tale che esso contenga un numero di osservazioni non inferiore ad una certa soglia fornita dal decision maker, identificano cluster con forma non convessa e sono in grado di isolare gli outlier

Attribuzioni

- Attribuzione esclusiva: ogni osservazione è assegnata ad un solo cluster ($\rightarrow K-$means)
- Attribuzione soft: ogni osservazione può appartenere a più cluster con diversi gradi di appartenenza

Categorie di clustering

- Metodi di partizione: suddividono il dataset in un numero definito di cluster non vuoti sulla base delle distanze tra le osservazioni ($\rightarrow K-$means)
- Metodi gerarchici: ricavano molteplici suddivisioni in cluster, sfruttando la struttura ad albero e utilizzando valori di soglia differenti
- Metodi basati sulla densità: sfruttano la densità locale per un intorno delle osservazioni, ragionano su un diametro specificato in modo tale che esso contenga un numero di osservazioni non inferiore ad una certa soglia fornita dal decision maker, identificano cluster con forma non convessa e sono in grado di isolare gli outlier

Attribuzioni

- Attribuzione esclusiva: ogni osservazione è assegnata ad un solo cluster ($\rightarrow K-$means)
- Attribuzione soft: ogni osservazione può appartenere a più cluster con diversi gradi di appartenenza
- Attribuzione completa: ogni osservazione viene assegnata ad almeno un cluster

Categorie di clustering

- Metodi di partizione: suddividono il dataset in un numero definito di cluster non vuoti sulla base delle distanze tra le osservazioni ($\rightarrow K-$means)
- Metodi gerarchici: ricavano molteplici suddivisioni in cluster, sfruttando la struttura ad albero e utilizzando valori di soglia differenti
- Metodi basati sulla densità: sfruttano la densità locale per un intorno delle osservazioni, ragionano su un diametro specificato in modo tale che esso contenga un numero di osservazioni non inferiore ad una certa soglia fornita dal decision maker, identificano cluster con forma non convessa e sono in grado di isolare gli outlier

Attribuzioni

- Attribuzione esclusiva: ogni osservazione è assegnata ad un solo cluster ($\rightarrow K-$means)
- Attribuzione soft: ogni osservazione può appartenere a più cluster con diversi gradi di appartenenza
- Attribuzione completa: ogni osservazione viene assegnata ad almeno un cluster
- Attribuzione parziale: alcune osservazioni possono essere non assegnate ad alcun cluster, molto utili per identificare presenza di outlier nel dataset