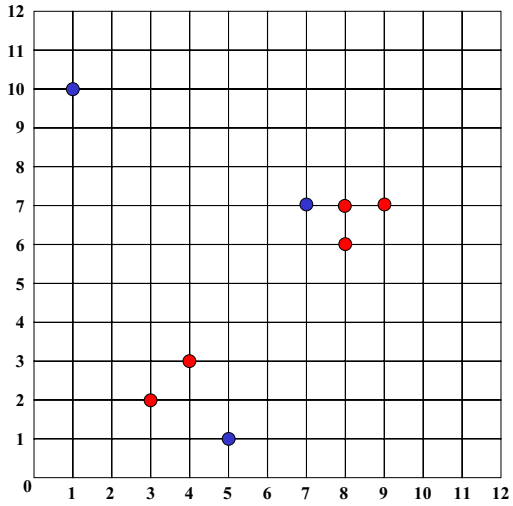
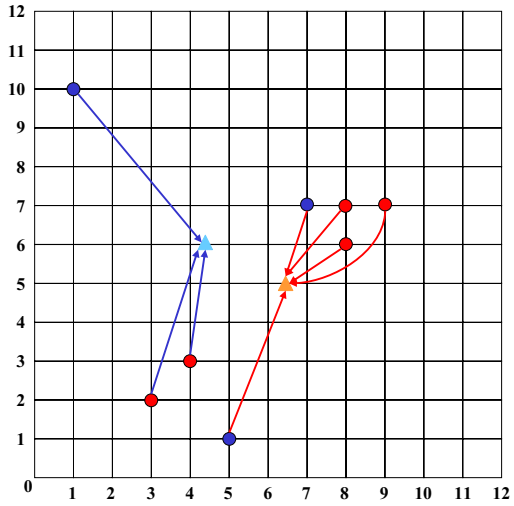
Clustering durch Varianzminimierung (k=2)



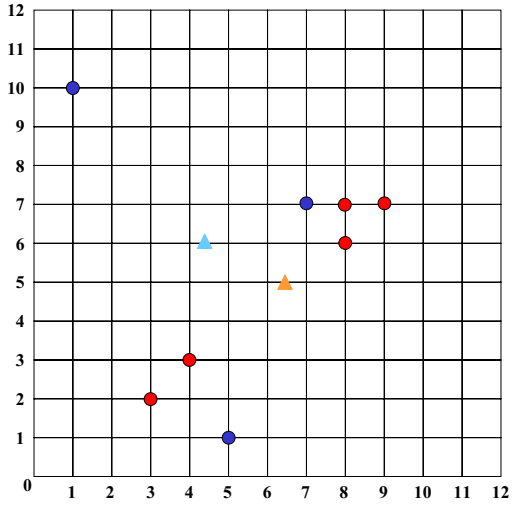
Initialisierung

Clustering durch Varianzminimierung (k=2)



Punkte neu zuordnen

Clustering durch Varianzminimierung (k=2)

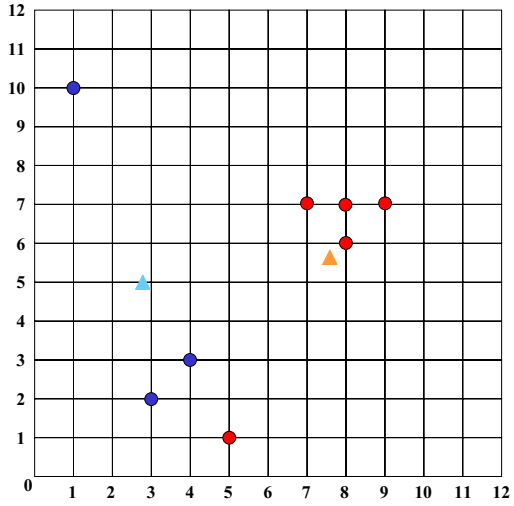


Zentroide berechnen

$\mu = (4.3 / 6.0)$

$\mu = (6.4 / 5.0)$

Clustering durch Varianzminimierung (k=2)

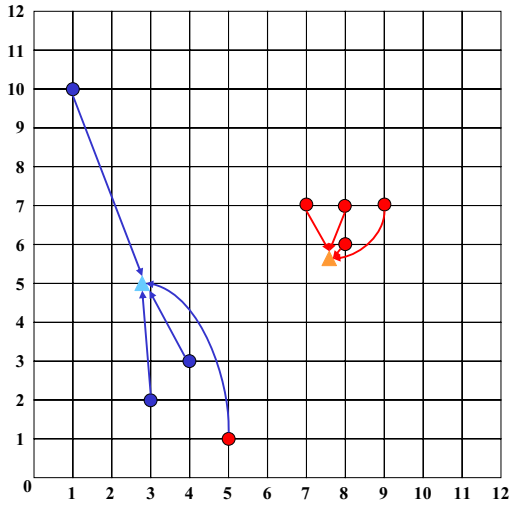


Zentroide berechnen

$\mu = (2.7 / 5.0)$

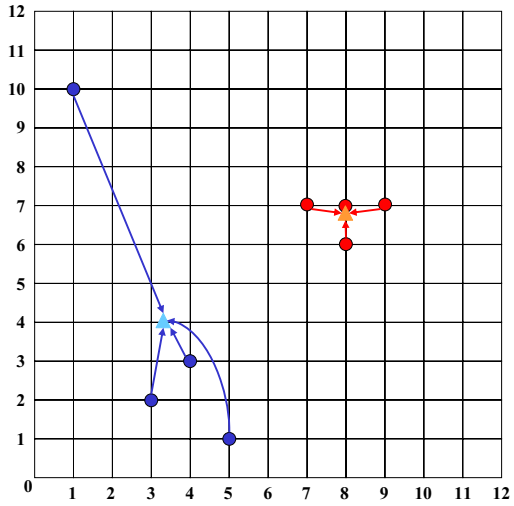
$\mu = (7.4 / 5.6)$

Clustering durch Varianzminimierung (k=2)



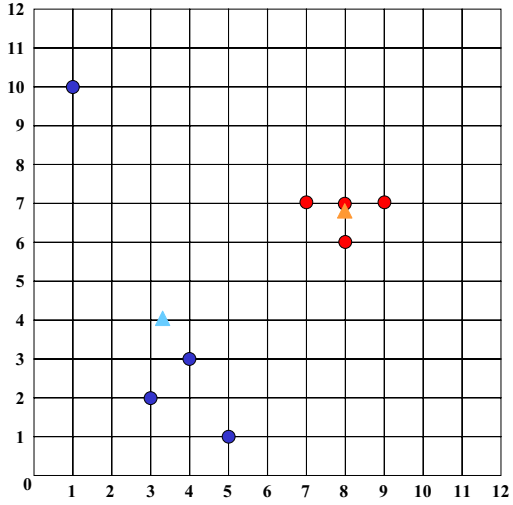
Punkte neu zuordnen

Clustering durch Varianzminimierung (k=2)



Punkte neu zuordnen

Clustering durch Varianzminimierung (k=2)

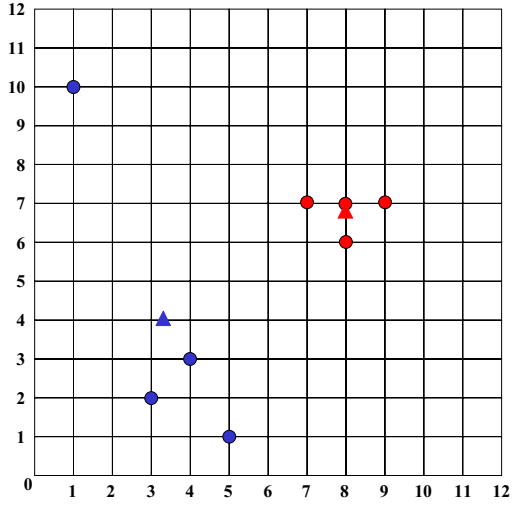


Zentroide berechnen

$\mu = (3.25 / 4.0)$

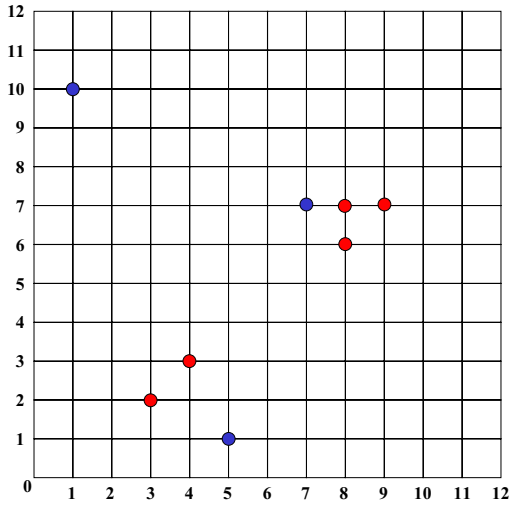
$\mu = (8.0 / 6.75)$

Clustering durch Varianzminimierung (k=2)



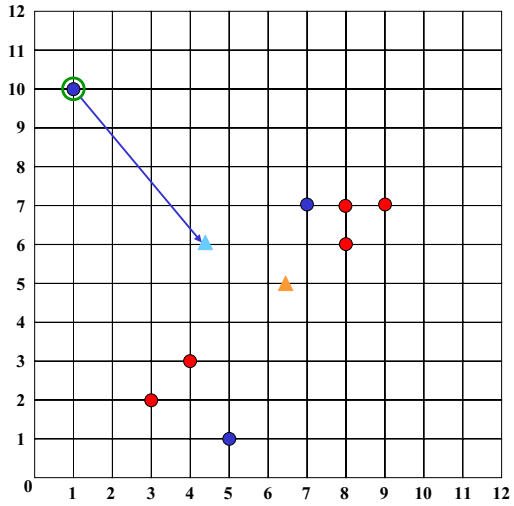
Konvergenz

k-means (k=2)



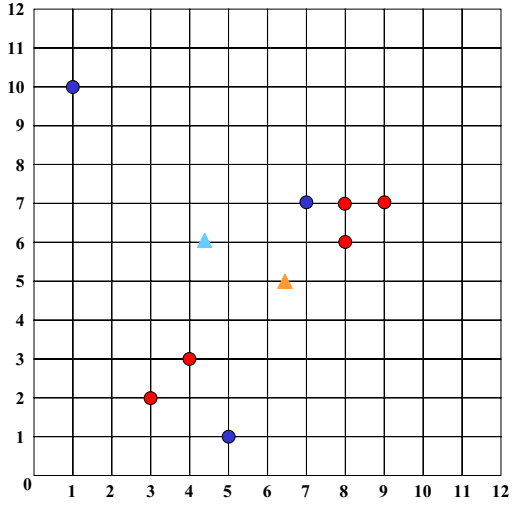
Initialisierung

k-means (k=2)



Zuordnung des ersten Punkts

k-means (k=2)

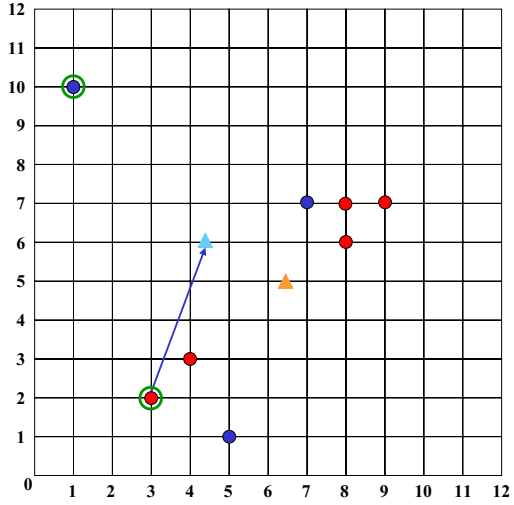


Zentroide berechnen

$\mu = (4.3 / 6.0)$

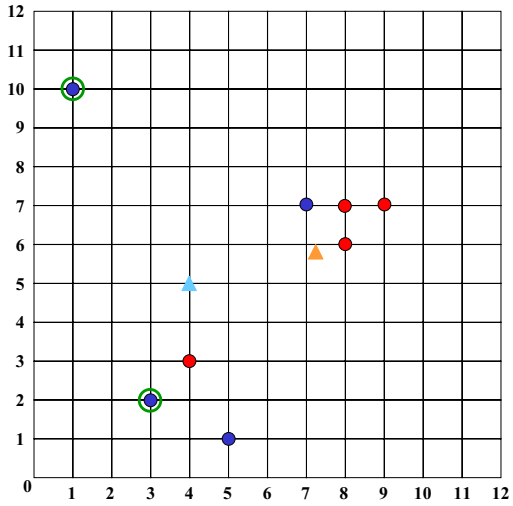
$\mu = (6.4 / 5.0)$

k-means (k=2)



Zuordnung des zweiten Punkts

k-means (k=2)

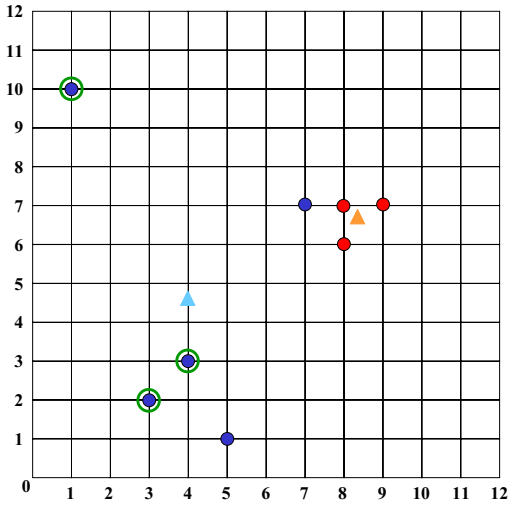


Zentroide berechnen

$\mu = (4.0 / 5.0)$

$\mu = (7.25 / 5.75)$

k-means (k=2)

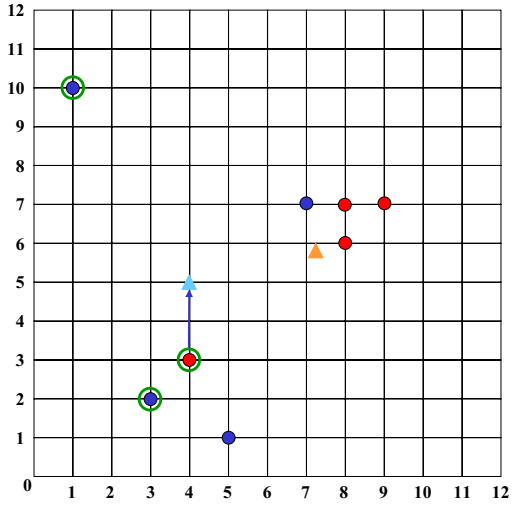


Zentroide berechnen

$\mu = (4.0 / 4.6)$

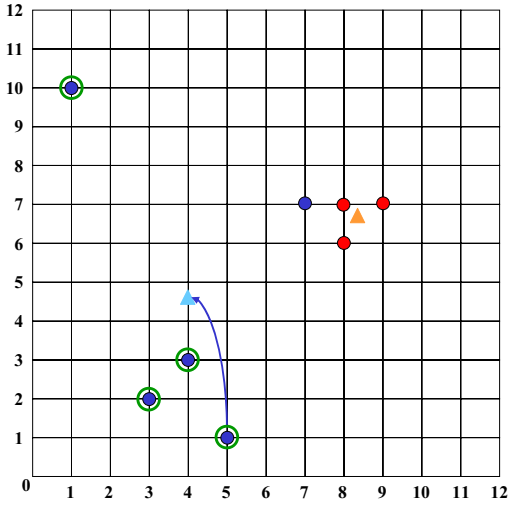
$\mu = (8.3 / 6.7)$

k-means (k=2)



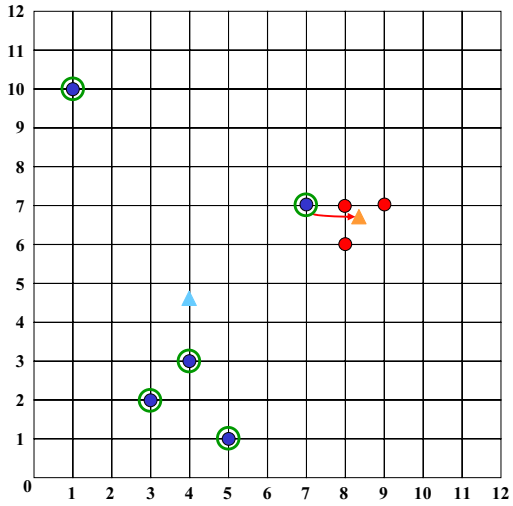
Zuordnen des dritten Punkts

k-means (k=2)



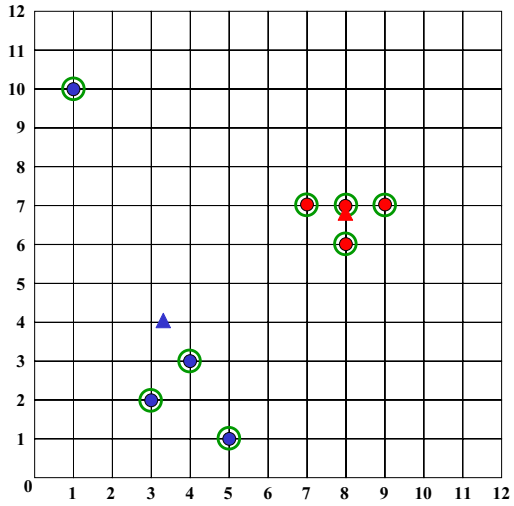
Zuordnung des vierten Punkts

k-means (k=2)



Zuordnung des fünften Punkts

k-means (k=2)

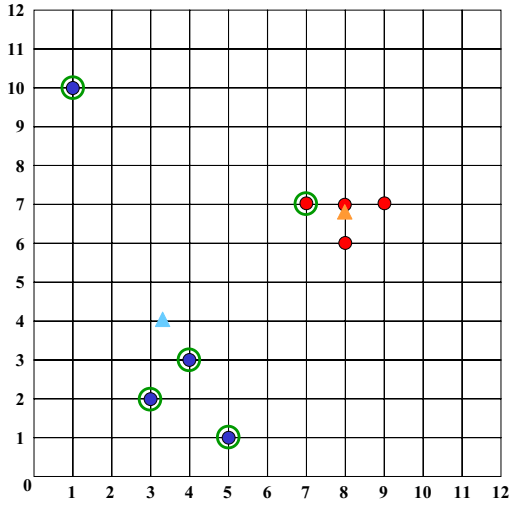


Keine weiteren Veränderungen => Konvergenz

$\mu = (3.25 / 4.0)$ ▲

$\mu = (8.0 / 6.75)$ ▲

k-means (k=2)

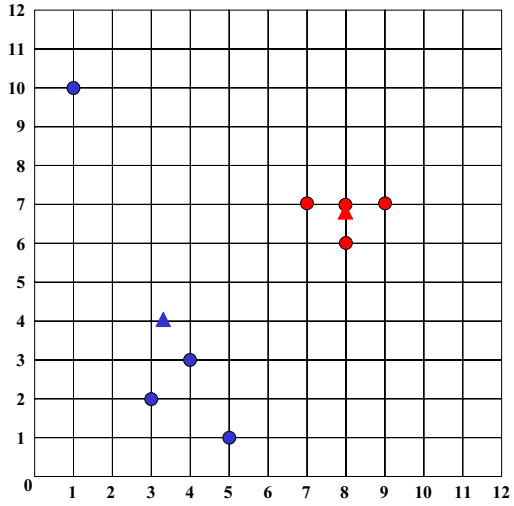


Zentroide berechnen

$\mu = (3.25 / 4.0)$ ▲

$\mu = (8.0 / 6.75)$ ▲

k-means (k=2)

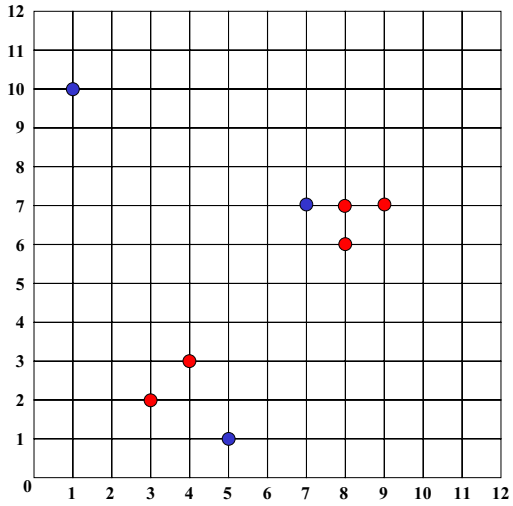


Ergebnis

$\mu = (3.25 / 4.0)$ ▲

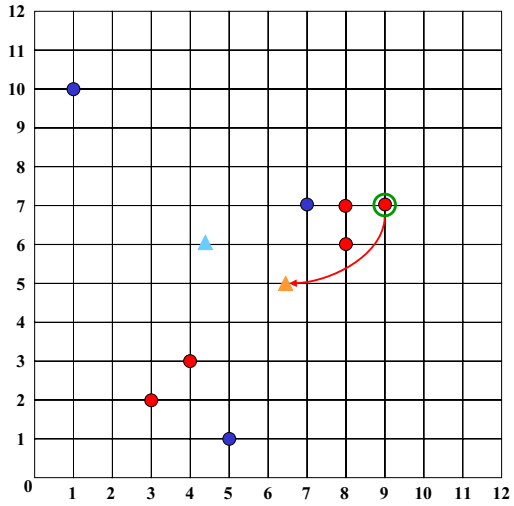
$\mu = (8.0 / 6.75)$ ▲

k-means (k=2) alternativer Ablauf



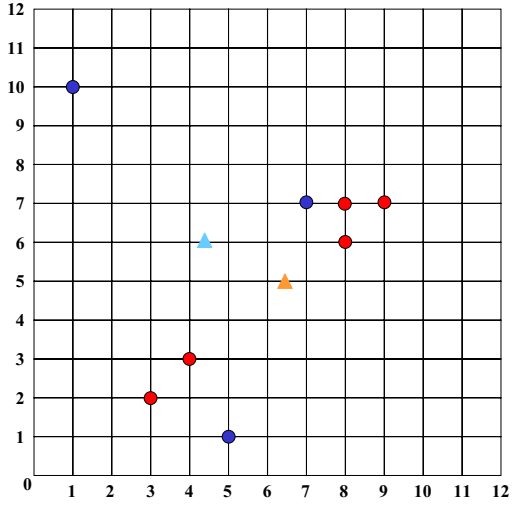
Initialisierung

k-means (k=2) alternativer Ablauf



Zuordnung des ersten Punkts

k-means (k=2) alternativer Ablauf

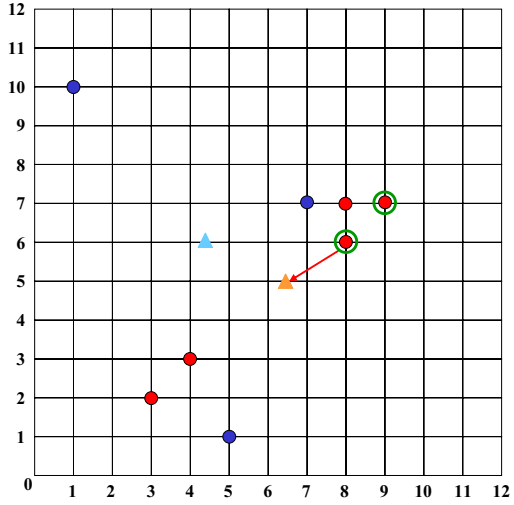


Zentroide berechnen

$\mu = (4.3 / 6.0)$

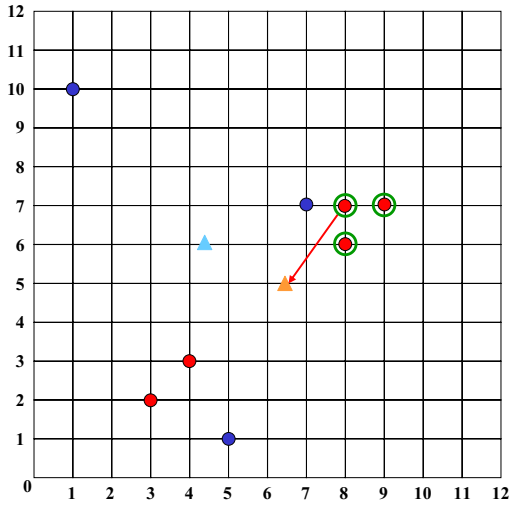
$\mu = (6.4 / 5.0)$

k-means (k=2) alternativer Ablauf



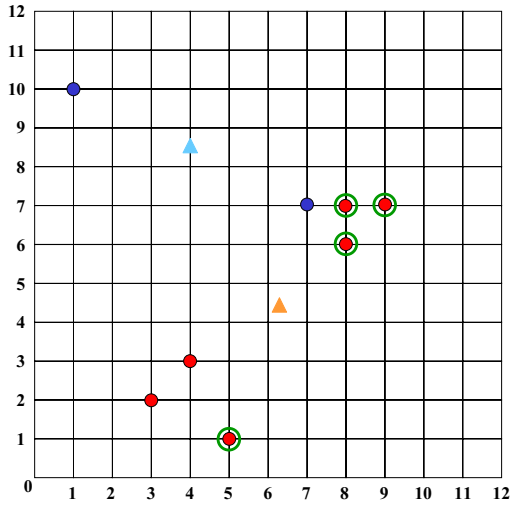
Zuordnung des zweiten Punkts

k-means (k=2) alternativer Ablauf



Zuordnung des dritten Punkts

k-means (k=2) alternativer Ablauf

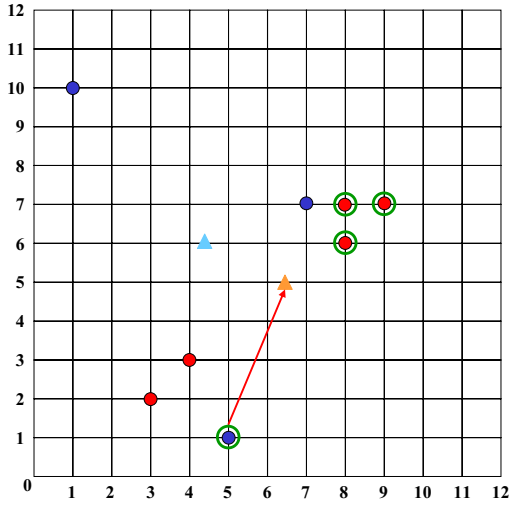


Zentroide berechnen

$\mu = (8.5 / 4.0)$

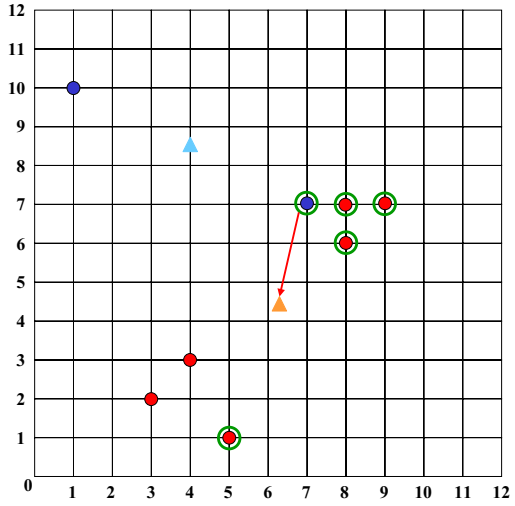
$\mu = (6.2 / 4.3)$

k-means (k=2) alternativer Ablauf



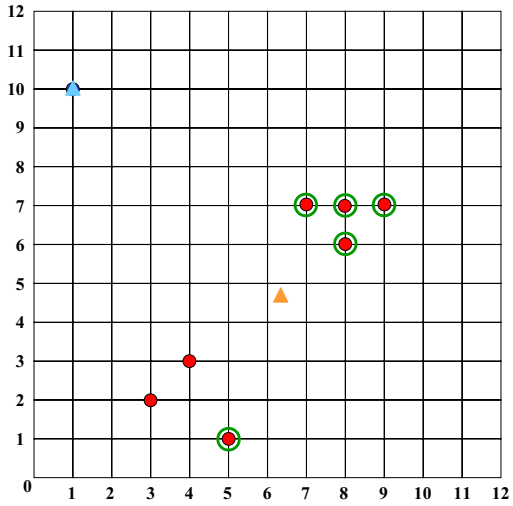
Zuordnung des vierten Punkts

k-means (k=2) alternativer Ablauf



Zuordnen des fünften Punkts

k-means (k=2) alternativer Ablauf



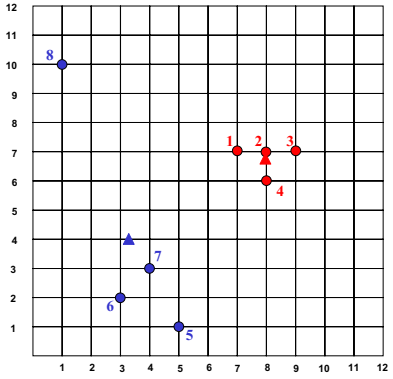
Zentroide berechnen

μ = (1.0 / 10.0)

μ = (6.3 / 4.7)

Kosten ?

Erzeugte Clusterings



μ = (3.25 / 4.0)

μ = (8.0 / 6.75)

Dist(μ,1) = |8.0 – 7.0| + |6.75 – 7.0| = 1 + 0.25 = 1.25

Dist(μ,2) = |8.0 – 8.0| + |6.75 – 7.0| = 0.25

Dist(μ,4) = |8.0 – 9.0| + |6.75 – 7.0| = 1.25

Dist(μ,3) = |8.0 – 8.0| + |6.75 – 6.0| = 0.75

$TD^2(C1) = 3.75$

Dist(μ,5) = |3.25 – 5.0| + |4.0 – 1.0| = 1.75 + 3 = 4.75

Dist(μ,6) = |3.25 – 3.0| + |4.0 – 2.0| = 0.25 + 2 = 2.25

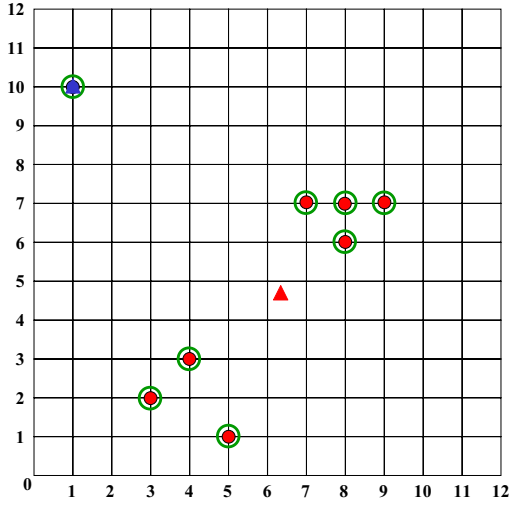
Dist(μ,7) = |3.25 – 4.0| + |4.0 – 3.0| = 0.75 + 1 = 1.75

Dist(μ,8) = |3.25 – 1.0| + |4.0 – 10.0| = 2.75 + 6 = 8.75

$TD^2(C2) = 107.25$

$TD^2 = 107.25 + 3.75 = 111$

k-means (k=2) alternativer Ablauf



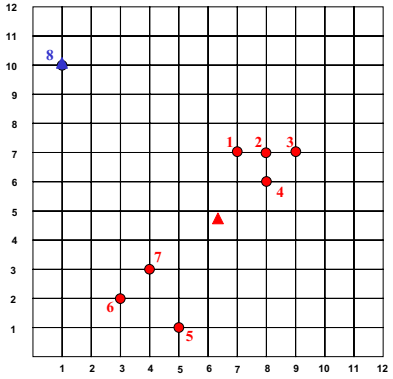
Keine weiteren Veränderungen => Konvergenz

μ = (1.0 / 10.0)

μ = (6.3 / 4.7)

Kosten ?

Alternatives Clusterings



μ = (1.0 / 10.0)

μ = (6.3 / 4.7)

Dist(μ,1) = |6.3 – 7.0| + |4.7 – 7.0| = 0.7 + 2.3 = 3.0

Dist(μ,2) = |6.3 – 8.0| + |4.7 – 7.0| = 1.7 + 2.3 = 4.0

Dist(μ,3) = |6.3 – 9.0| + |4.7 – 7.0| = 2.7 + 2.3 = 5.0

Dist(μ,4) = |6.3 – 8.0| + |4.7 – 6.0| = 1.7 + 1.3 = 3.0

Dist(μ,5) = |6.3 – 5.0| + |4.7 – 1.0| = 1.3 + 3.7 = 5.0

Dist(μ,6) = |6.3 – 3.0| + |4.7 – 2.0| = 3.3 + 2.7 = 6.0

Dist(μ,7) = |6.3 – 4.0| + |4.7 – 3.0| = 2.3 + 1.7 = 4.0

$TD^2(C1) = 136$

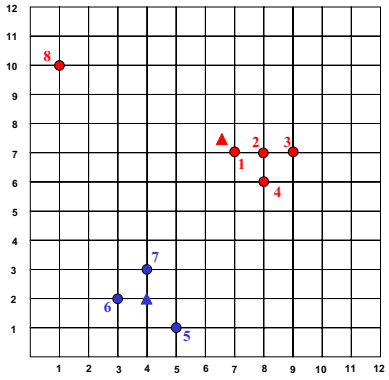
Dist(μ,8) = |1.0 – 1.0| + |10.0 – 10.0| = 0

$TD^2(C2) = 0$

$TD^2 = 136$

Kosten ?

Optimales Clustering



$\mu = (4.0 / 2.0)$

$\mu = (6.6 / 7.4)$

$Dist(\mu,1) = |6.6 - 7.0| + |7.4 - 7.0| = 0.4 + 0.4 = 0.8$

$Dist(\mu,2) = |6.6 - 8.0| + |7.4 - 7.0| = 1.4 + 0.4 = 1.8$

$Dist(\mu,3) = |6.6 - 8.0| + |7.4 - 6.0| = 1.4 + 1.4 = 2.8$

$Dist(\mu,4) = |6.6 - 9.0| + |7.4 - 7.0| = 2.4 + 0.4 = 2.8$

$Dist(\mu,8) = |6.6 - 1.0| + |7.4 - 10.0| = 5.6 + 2.6 = 8.2$

$TD^2(C1) = 86.8$

$Dist(\mu,5) = |4.0 - 5.0| + |2.0 - 1.0| = 2$

$Dist(\mu,6) = |4.0 - 3.0| + |2.0 - 2.0| = 1$

$Dist(\mu,7) = |4.0 - 4.0| + |2.0 - 3.0| = 1$

$TD^2(C2) = 6$

$TD^2 = 86.8 + 6.0 = 92.8$

Schluss

- Merke:
  - K-means konvergiert meist nur gegen lokales Minimum
  - K-means ist abhängig von der Reihenfolge der Zuordnung und der initialen Partitionierung
  - K-means ist anfällig gegen Rauschen
  - K-means ist trotzdem das zur Zeit populärste Clustering Verfahren in der Industrie und auch in weiten Teilen der Wissenschaft.