

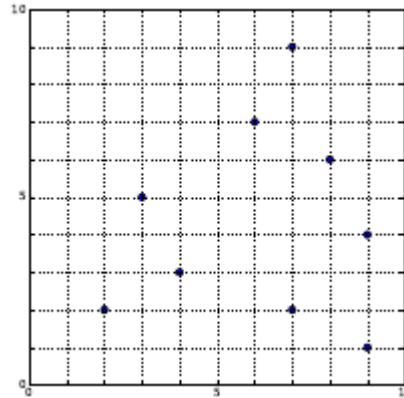
Blatt 11 (8 Punkte + 1 Bonuspunkt)

Abgabe durch Hochladen auf der eCampus-Seite bis Sonntag, 03.07.2016, 23:59 Uhr, in Gruppen von 2-3 Personen.

Aufgabe 11.1: k-Means

(2)

Das folgende Bild zeigt Datenpunkte, die bezüglich zwei Attributen dargestellt sind:



Wenden Sie den k-Means Algorithmus an. Initialisieren Sie die Schwerpunkte mit den Werten $m_1^{(0)} = (2, 2)$, $m_2^{(0)} = (7, 9)$ und $m_3^{(0)} = (9, 4)$. Ordnen Sie für jede Iteration die Datenpunkte ihren zugehörigen Clusterzentren zu und geben Sie diese Zuordnung explizit an. Berechnen Sie dann die neuen Schwerpunkte $m_i^{(t+1)}$. Geben Sie die Herleitung der Schwerpunkte wieder. Iterieren Sie solange, bis die Cluster stabil sind, d.h. keine Neuzuweisung von Datenpunkten zu Schwerpunkten mehr vorgenommen wird.

Aufgabe 11.2: Clustering über Single-, Complete-Linkage, DBSCAN, OPTICS

(0.5 + 1.5 + 1.5 + 0.5 + 0.5(Bonus) + 0.5(Bonus) = 4 + 1Bonus)

Gegeben sind die folgenden 2-dimensionalen Datenpunkte:

$$a = (1, 2), b = (3, 1), c = (2, 4), d = (6, 5), e = (8, 4), f = (1, 6).$$

- a) Vervollständigen Sie die folgende Tabelle mit den Distanzen zwischen den Punkten. Es reicht hierbei aus, nur die Werte oberhalb der Diagonale einzutragen.

	a	b	c	d	e	f
a	0	$\sqrt{5}$	$\sqrt{5}$	$\sqrt{34}$	$\sqrt{53}$	$\sqrt{16}$
b		0				
c			0			
d				0		
e					0	
f						0

- b) Wenden Sie den Algorithmus für das AGGLOMERATIVE CLUSTERING durch SINGLE-LINKAGE an. Hierbei soll der Algorithmus terminieren, wenn eine Distanz von $\sqrt{5}$ überschritten würde. Die Lösung muss dabei nach jedem Agglomerationsschritt die jeweils neue Distanztabelle zeigen. Die Werte in den neuen Distanztabellen dürfen wie in der der Ausgangstabelle unausgerechnete Wurzeln von Zahlen enthalten.

- c) Welches Ergebnis liefert der gleiche Algorithmus mit COMPLETE-LINKAGE und einer Distanzgrenze von $\sqrt{10}$? Wieder muss die Lösung dabei nach jedem Agglomerations-schritt die jeweils neue Distanztabelle zeigen. Die Werte in den neuen Distanztabellen dürfen wie in der Ausgangstabelle unausgerechnete Wurzeln von Zahlen enthalten. Ist das Ergebnis eindeutig? Begründen Sie Ihre Antwort bzgl. der Eindeutigkeit!
- d) Welches sind die Kernobjekte bei den oben gegebenen Datenpunkten und DBSCAN mit der Parameterbelegung $\text{minNeighbors} = 2, \epsilon = \sqrt{5}$?
- e) **Bonus-Aufgabe:** Berechnen Sie die OPTICS Kerndistanz für jeden Punkt mit der Parameterbelegung $\text{minNeighbors} = 2, \epsilon = \sqrt{5}$.
- f) **Bonus-Aufgabe:** Ist die Wahl der Euklidischen Distanz bei periodischen Größen, wie Wochentagen, sinnvoll? Begründen Sie Ihre Antwort.

Aufgabe 11.3: Lineare Grauwertspreizung und Mittelwertfilter (1 + 0.5 + 0.5 = 2)

Grundsätzlich: Für die Umwandlung von Zahlen mit Nachkommastellen in ganze Zahlen verwenden Sie Rundung!

184	120	72	56
160	140	64	60
128	112	96	64
120	102	88	80

Obige 4×4 Matrix stellt einen Ausschnitt der Grauwerte eines Bildes dar.

- a) Maximieren Sie den Kontrast dieses Ausschnitts für einen möglichen Wertebereich von $\{0, 1, \dots, 255\}$ durch lineare Spreizung (Vorlesung 19). Geben Sie die neu berechneten Werte wieder in einem 4×4 großen Feld an.
- b) Wenden Sie auf die oberen zwei zentralen Pixel des Ausschnittes (hervorgehobene obere Pixel) ein 3×3 -Pixel großes Mittelwertfilter (Vorlesung 19) an.
- c) Geben Sie eine plausible Lösung an, wie das Mittelwertfilter zur Anwendung auf die Randpixel angepasst werden sollte. Wenden Sie das angepasste Mittelwertfilter auf das hervorgehobene Pixel am Rand an und begründen Sie die von Ihnen ausgewählte Anpassungsmöglichkeit.