

Online-Hausarbeit 5

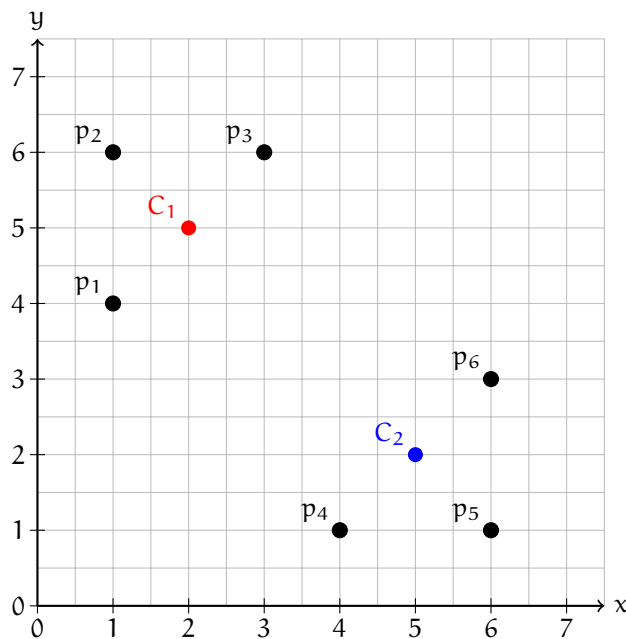
Intelligente Systeme im Sommersemester 2020

Abgabetermin: Geben Sie Ihre Lösung im Uni2Work bis zur Abgabefrist am 17.07.2020, 18:59 Uhr, ab! Sollten Sie nachweislich Internetprobleme haben, die eine Abgabe bis 18:59 Uhr nicht ermöglichen, so geben Sie bitte bis 23:59 Uhr ab und schreiben uns parallel dazu eine E-Mail, in der Sie um eine verlängerte Abgabefrist bitten und Ihre Umstände erklären.

Aufgabe 1: Fuzzy-Clustering

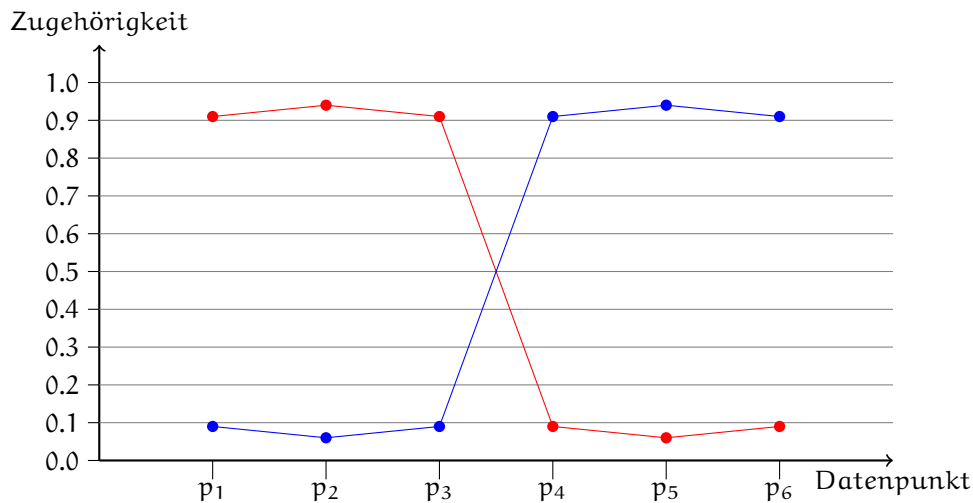
(3,5+1,5 Pkt.)

Gegeben sei ein Datensatz von sechs Datenpunkten p_i mit $i \in [1, 6]$ sowie zwei Clusterzentren C_1 und C_2 :



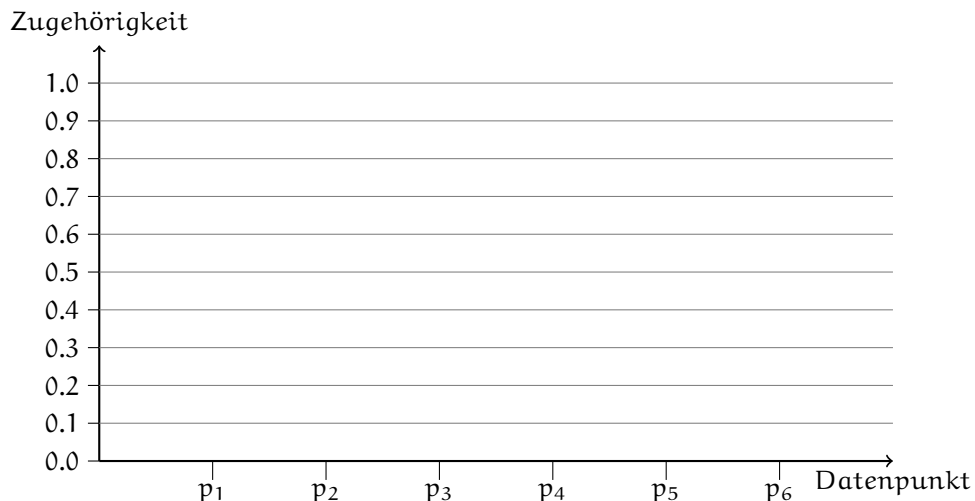
Folgende Grafik zeigt das Ausführungsergebnis eines Fuzzy-Clusterings, so wie Sie es in der Vorlesung auf Seite 272 kennen gelernt haben. Sie sehen die optimalen Zugehörigkeiten der Punkte zu den Clustern für einen Fuzzifier $w = 2$. Zugehörigkeiten zu Cluster C_1 sind mit roten, Zugehörigkeiten zu Cluster C_2 mit blauen Punkten markiert.

Hinweis: Das Verbinden der Punkte je Cluster mittels Linien ist semantisch natürlich nicht korrekt. Diese Hilfslinien wurde jedoch trotzdem eingezeichnet, damit die relativen Unterschiede der Werte besser sichtbar sind. Verwenden Sie im Folgenden bitte ebenfalls geeignet verbundene Punkte.

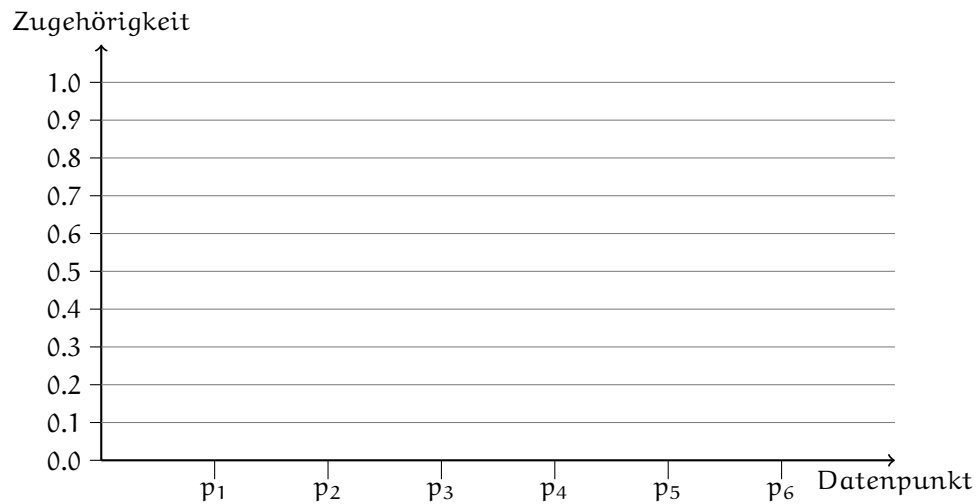


- a. Zeichnen Sie analog zur obigen Abbildung jeweils einen Plot für $w = 1$ und $w = 5$.
Hinweis: Im oben gegebenen Plot für $w = 2$ sind die berechneten optimalen Zugehörigkeitswerte dargestellt. Sie müssen in dieser Aufgabe jedoch nicht rechnen. Wenn Sie "fuzzy" Zugehörigkeitswerte einzeichnen, reicht es aus, wenn die Auswirkung des Fuzzifiers auf die Zugehörigkeiten deutlich wird. Wichtig ist die relative Verschiebung der Zugehörigkeiten. Verwenden Sie ebenfalls die Farben rot und blau oder markieren Sie die Zugehörigkeitswerte anderweitig.

(i) Plot für $w = 1$:



(ii) Plot für $w = 5$:



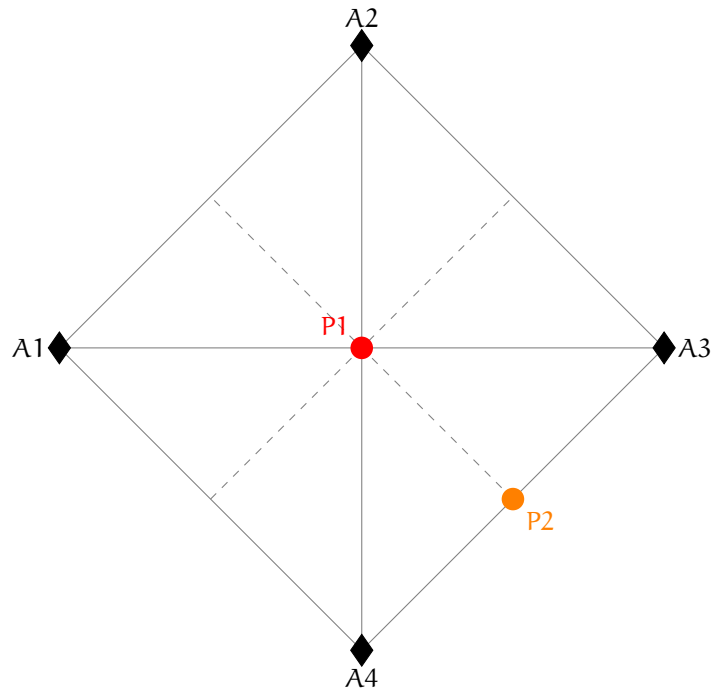
- b. Bei der Verwendung eines Fuzzifiers $w > 1$ ist es in den meisten Fällen sinnvoll, einen Datenpunkt durch "fuzzy" statt harten Zugehörigkeiten zu Clustern zuzuordnen. Wo muss ein Punkt liegen, sodass es auch für $w > 1$ besser ist, diesen Punkt durch harte statt "fuzzy" Zugehörigkeiten zu Clustern zuzuordnen? Begründen Sie dies.

Geben Sie für obiges Beispiel außerdem an, welche Koordinaten ein Punkt haben muss, damit dieser eine Zugehörigkeit von 1 zu einem der beiden Clusterzentren besitzt.

Aufgabe 2: Archetypische Analyse

(3+1,5+2,5 Pkt.)

- a. Gegeben sei ein im Koeffizientenraum dargestellter Datensatz mit eingezeichneten Hilfslinien zur Zugehörigkeitsgewichtung. Ferner sind vier Archetypen A1, A2, A3 und A4 sowie zwei Punkte P1 und P2 markiert.



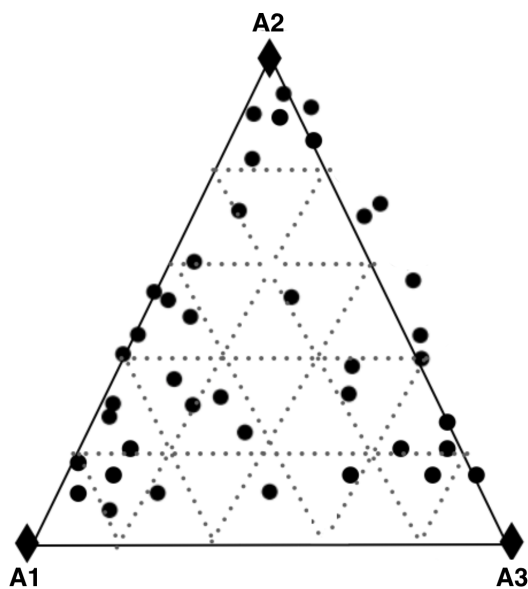
- (i) Betrachten Sie Punkt P1. Können Sie basierend auf obiger Darstellung entscheiden, ob P1 eine einzelne Zugehörigkeitskonstellation zu den Archetypen besitzt ("eindeutig") oder ob mehrere Konstellationen möglich sind ("mehrdeutig")? Wenn Sie denken, dass die Konstellation an Zugehörigkeiten zu den Archetypen eindeutig ist, dann tragen Sie diese in folgende Tabelle ein. Wenn Sie denken, dass die Konstellation an Zugehörigkeiten mehrdeutig ist, dann begründen Sie dies anhand von Beispielen.

	A1	A2	A3	A4
P1				

- (ii) Betrachten Sie nun Punkt P2. Können Sie basierend auf obiger Darstellung entscheiden, ob die Zugehörigkeitskonstellation von P2 wie oben definiert eindeutig oder mehrdeutig ist? Wenn Sie denken, dass die Konstellation an Zugehörigkeiten zu den Archetypen eindeutig ist, dann tragen Sie diese in folgende Tabelle ein. Wenn Sie denken, dass die Konstellation an Zugehörigkeiten mehrdeutig ist, dann begründen Sie dies anhand von Beispielen.

	A1	A2	A3	A4
P2				

- b. Gegeben sei nun ein neuer Datensatz, erneut dargestellt im Koeffizientenraum mit eingezeichneten Hilfslinien zur Zugehörigkeitsgewichtung sowie drei Archetypen A1, A2 und A3.



Gegeben sei außerdem folgende Tabelle, aus der Sie die Fuzzy-Zugehörigkeiten der drei Punkte P1, P2 und P3 zu den Archetypen A1, A2 und A3 auslesen können.

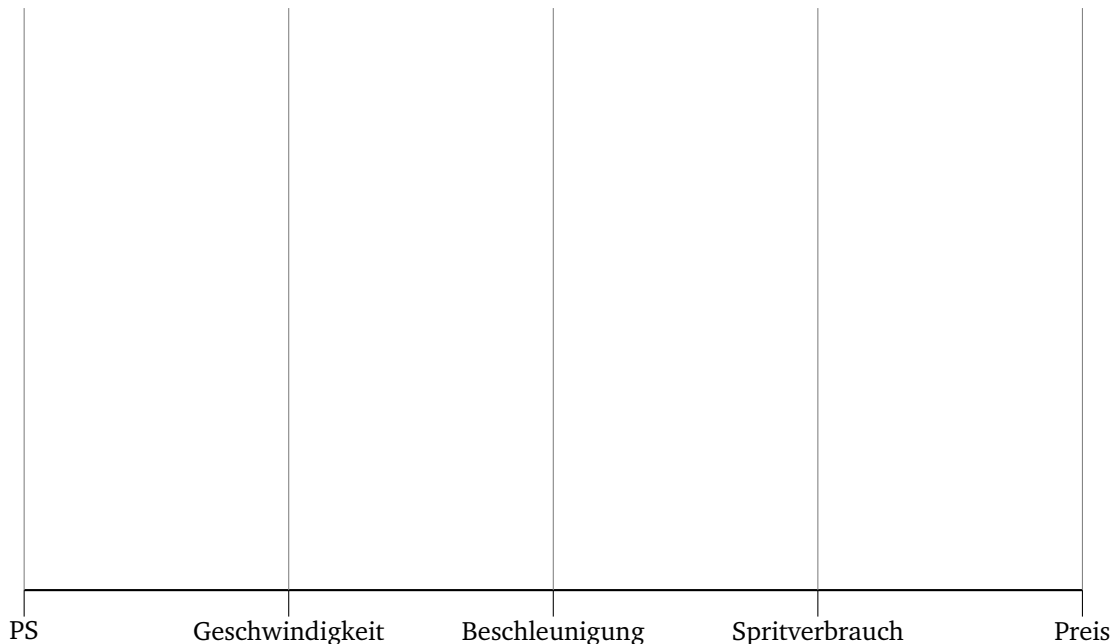
	A1	A2	A3
P1	0,4	0,4	0,2
P2	0	0,6	0,4
P3	0,2	0,2	0,6

Zeichnen Sie die drei Punkte in die oben abgebildete Grafik ein und markieren Sie die Punkte mit P1, P2 und P3.

- c. In folgender Tabelle sehen Sie einen Datensatz, in dem zehn Autos x_i für $i \in [0, 9]$ anhand der fünf Eigenschaften *PS*, *max. Geschwindigkeit in km/h*, *Beschleunigung von 0 auf 100 km/h in Sekunden*, *Spritverbrauch in l/100km* und dem *Preis in Euro* verglichen werden.

	PS	Geschwindigkeit	Beschleunigung	Spritverbrauch	Preis
x_0	120	170	9,0	8,5	30.000
x_1	100	170	9,5	9,0	25.000
x_2	120	180	9,1	10,5	35.000
x_3	200	250	7,2	10,5	60.000
x_4	180	220	8,0	11,5	50.000
x_5	230	300	7,3	10,5	55.000
x_6	250	300	7,0	12,0	75.000
x_7	170	220	8,0	9,5	55.000
x_8	210	270	7,3	10,0	60.000
x_9	190	250	8,1	9,0	55.000

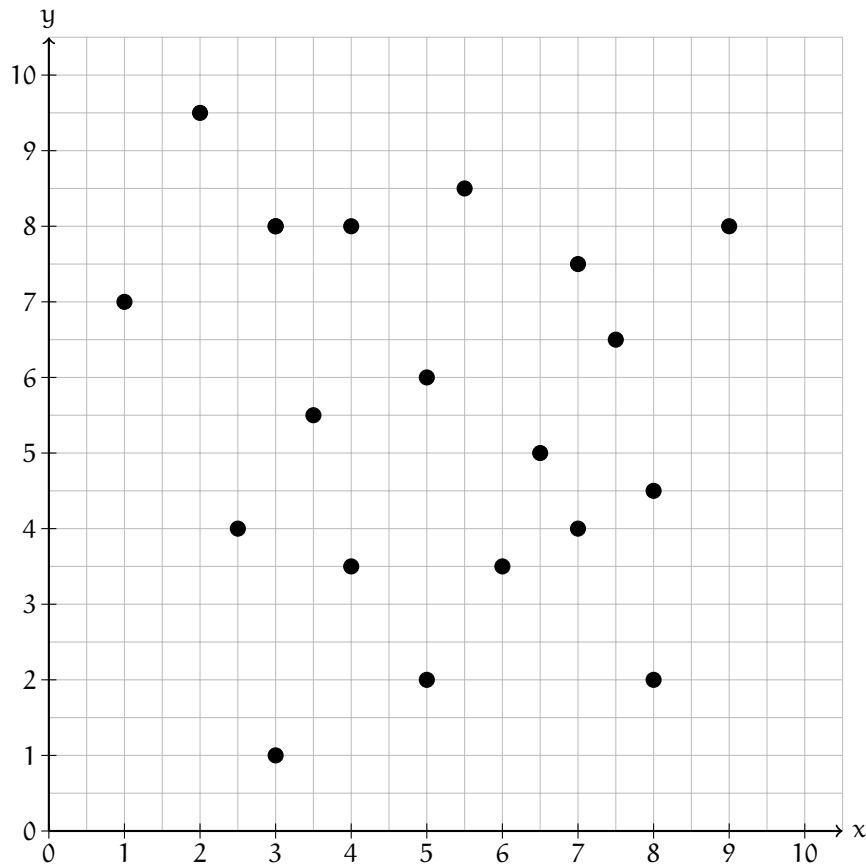
Ergänzen Sie nun das folgende Parallelkoordinatensystem mit drei unterschiedlichen Datenpunkte aus der Tabelle. Die drei letzten unterschiedlichen Ziffern Ihrer Matrikelnummer sollen die drei Indizes der zu wählenden Datenpunkte darstellen. Wenn Ihre Matrikelnummer auf 1337 endet, dann zeichnen Sie entsprechend die Datenpunkte x_1 , x_3 und x_7 ein. Wählen Sie außerdem für jede Achse eine geeignete Skalierung und Beschriftung. Vergessen Sie nicht, auch die eingezeichneten Datenpunkte zu beschriften.



Aufgabe 3: Archetypische Analyse – Residuenquadratsumme

(3 Pkt.)

Gegeben sei folgender Datensatz:



Zunächst müssen Sie den Datensatz personalisieren, indem Sie einen weiteren Punkt mit den Koordinaten $(10, m + 1)$ hinzufügen. Hierbei steht m für die letzte Ziffer Ihrer Matrikelnummer. Wenn Ihre Matrikelnummer also auf $m = 2$ endet, so besitzt der neu einzuzeichnende Punkt die Koordinaten $(10, 3)$. Zeichnen Sie den Punkt in obige Abbildung ein und bearbeiten Sie folgende Aufgabe **für Ihren personalisierten Datensatz**.

Wie Sie aus der Vorlesung wissen, ist das Ziel bei der archetypischen Analyse, die Residuenquadratsumme (RSS) zu minimieren. Wie viele Archetypen werden für Ihren personalisierten Datensatz **mindestens** benötigt, sodass $RSS = 0$? Begründen Sie Ihre Antwort und zeichnen Sie die benötigten Archetypen in Ihren personalisierten Datensatz ein.

Selbstständigkeitserklärung

Hiermit versichere ich, dass die abgegebene Lösung alleinig durch mich angefertigt wurde und ohne die Hilfe Dritter entstanden ist. Insbesondere habe ich keine Lösungen von Dritten teilweise oder gänzlich abgegeben.

Matrikelnummer, Name

Ort, Datum

Unterschrift