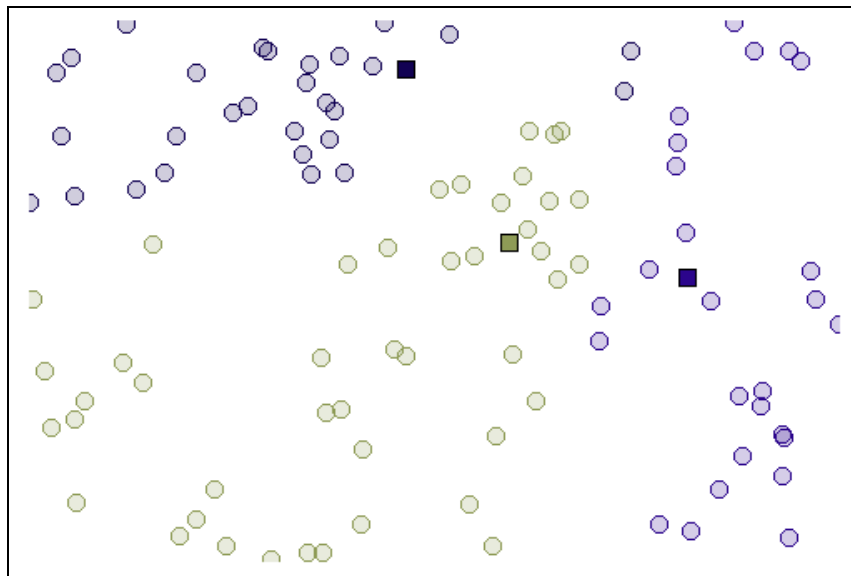


Clustering

1. In einem Zwischenschritt des k-Means-Algorithmus mit $k=3$ ist das im Bild dargestellte Zustand entstanden. Skizzieren Sie die Änderungen welche in dem nächsten Schritt des Algorithmus erfolgen.



2. Die Dokumente, die für ein und denselben Konzept (z.B. Auto) unterschiedliche Terme benutzen (e.g. Pkw, VW Golf, etc.) werden beim Clustering mit einer hohen Wahrscheinlichkeit ein und demselben Cluster zugeordnet. Erklären Sie warum.
3. Erklären Sie die Gemeinsamkeiten und Unterschiede zwischen Klassifikation und Clustering. Achten Sie insbesondere auf die folgenden Eigenschaften:
 - a) Training des Systems
 - b) Anzahl der Cluster/Klassen
 - c) Erklärbarkeit und Darstellung des Ergebnisses