Allgemeines

Da die Testversion meines Latexeditor unerwartet kurz vor der Abgabe geendet ist, sind folgen die letzten Zusätze hier. Das ist auch der Grund dafür, dass die Bearbeitung von Aufgabe 1 keinen Quellcode enthält. Diesen kann man allerdings in kommentierter Form in den python-Dateien ansehen.
Ich entschuldige mich hierfür und bitte um Verständnis.

Die Programme sind in den jeweiligen main.py Dateien zu starten. Dort kann über den Parameter in der get_data Methode die Beispieldatei als Eingabe gewählt werden. Beide Programme nutzen nutzen aktuell den besten entwickelten Algorithmus. Dies kann aber ohne weiteres im Programm geändert werden.

Zusatz Aufgabe 1:

Abstract:

Die folgende Bearbeitung beschäftigt sich mit einer Version des TSP, das Quadratic Bounded Angular Metric TSP genannt werden kann. Hierbei meint Quadratic, dass die Kostenfunktion nicht nur von den letzten zwei, sondern den letzten drei besuchten Punkten abhängt. Bounded bedeutet, dass die Kosten einer Kante durch eine Obergrenze, hier der maximale Abbiegewinkel von 90° begrenzt sind. Und Angular Metric bedeutet, dass hier der Abbiegewinkel an den Punkten als Metrik betrachtet wird.
Es handelt sich um eine Klasse des TSP, die noch nicht in der Literatur zu finden ist. Deshalb wird die NP-Vollständigkeit zunächst bewiesen. Dann wird zunächst ein exaktes ineffektives Lösungsverfahren entwickelt, dass dann durch einige Heuristiken verbessert wird und zur Lösung der Beispielaufgaben verwendet wird. Zum Schluss wird das zugehörige Entscheidungsproblem betrachtet und eine Art Metaheuristik entwickelt, die das Finden einer Lösung deutlich beschleunigt, aber nicht exakt ist. Das heißt, dass nicht garantiert ist, dass eine Lösung gefunden wird, obwohl eine existiert. Außerdem kann die Matrix genutzt werden, um schwer integrierbare Punkte zu bestimmen (Punkte, die wenig Abbiegewinkel <=90° haben sind schwer integrierbar). Dies kann für eine Periodisierung dieser Punkte als Startpunkt genutzt werden.

Zusatz zu Kapitel 5.2: Ich vermute, dass es keine Punkte geben kann, die nicht Teil der konvexen Hülle der Punktemenge sind und einen minimalen Abbiegewinkel >90° haben. Den Beweis bleibe ich allerdings schuldig, weshalb die Generalisierung nicht überflüssig ist. Außerdem kann die Anzahl

Die Nearest-Neighbour Heuristik wird in der Bearbeitung öfter mit NN-Heuristik abgekürzt. Diese Abkürzung wurde nicht eingeführt.

Die NP-Schwere für das HAMILTON-Problem ist im Artikel ``EXPECTED COMPUTATION TIME FOR HAMILTONIAN PATH PROBLEM`` von Yuri Urevich bewiesen (https://shelah.logic.at/files/95591/243.pdf).

Der Quickhull-Algorithmus wird im Artikel ``The Quickhull Algorithm for Convex Hulls`` von David Dobkin vorgestellt (https://dl.acm.org/doi/pdf/10.1145/235815.235821).

Zusatz Aufgabe 2:

Verbesserung des kmeans Verfahrens: Wenn die Peaks im Differenzplot zusammenschmelzen, lassen sich die Daten nicht mehr sauber Clustern. Um dies zu erkennen kann der silhouette_score gebildet werden. Da dies nur in O(n^2) möglich ist, ist diese Methode aber sehr ineffizient. Da die Peaks durch die stochastische Verteilung der Scheiben mit der Wahrscheinlichkeit 1/3 zustande kommt, enthalten alle Peaks ungefähr die selbe Menge von Elementen. Man könnte anhand des

Indexes und der Länge der Liste mit den Differenzen ungefähr ein Drittel der Liste abtrennen und so eine Abschätzung der Cluster erhalten. Jetzt kann anhand von Lagemaßem (Mittelwert, Standardabweichung) die Verteilung der Daten innerhalb der Liste analysiert werden und Rückschlüsse darauf gezogen werden, wie gut sich die Daten Clustern lassen.

Standardabweichung ungefähr so groß, wie das Intervall, in dem die Daten liegen -> Daten sind recht gleichmäßig in den Clustern verteilt und die Daten eignen sich eher für das “second” Verfahren.

Standardabweichung wesentlich kleiner, als das Intervall, in dem die Daten liegen -> Daten bilden klar definierte Peaks und das “kmeans” Verfahren ist geeignet.