

1. Übungszettel Mustererkennung WS15/16

Prof. Raúl Rojas, Fritz Ulbrich
Institut für Informatik, Freie Universität Berlin
Abgabe Online bis Mittwoch, 04.11.15, 10 Uhr

1. Aufgabe (2 Punkte): Intro

- lade die MNIST-Datensätze von <http://yann.lecun.com/exdb/mnist/> herunter
- schreibe ein Skript zum Laden der Daten
- schreibe ein Skript zum Anzeigen von 10 zufällig ausgewählten Datenpunkten und ihrem Label
- gib die Ausgabe vom vorletzten Punkt als Lösung der Aufgabe 1 ab

2. Aufgabe (4 Punkte): k-NN

- implementiere k-Nearest Neighbor, wie in der Vorlesung besprochen
- benutze Deinen Algorithmus und die Trainingsdaten zum Klassifizieren der Testdaten für $k = 1, 3, 11$
- berichte tabellarisch wieviel Prozent der Daten richtig klassifiziert werden konnten

3. Aufgabe (4 Punkte): k-means

- implementiere k-means, wie in der Vorlesung besprochen
- benutze Deinen Algorithmus, um im Trainingsdatensatz Cluster zu finden für $k = 9, 10, 20$
- zeige für jedes dieser k 10 zufällige Samples jedes gefundenen Clusters an
- was wäre ein sinnvolles Fehlermaß für ein Clusteringergebnis? Warum?
- berechne es!