

Programmierabgabe - Gruppe 9

Klassifikation von Hunderassen mit Scikit Learn

Datenset

Das Datenset "Stanford Dogs" ist ein Subset des ImageNet-Datensets und besteht aus Aufnahmen von 120 verschiedenen Hunderassen. Pro Rasse stehen dabei 150 - 250 Samples zur Verfügung. Sie können das Datenset unter <http://vision.stanford.edu/aditya86/ImageNetDogs/> downloaden. Dort finden sich zwar bereits Listen mit Trainings- und Testsplits sowie Merkmalsmatrizen für Training und Test, aber laden Sie nur Bilder und Labels (Annotations) herunter.

Für die Bearbeitung der weiteren Aufgabe suchen Sie sich bitte nur 5 der Hunderassen aus und verwenden Sie nur diese Bilder für Ihre Untersuchungen. Geben Sie in Ihrem Kurzreport bitte an, welche Rassen Sie gewählt haben und warum sie sich für diese entschieden haben. Die Bilder der jeweiligen Hunderassen können Sie auf der Website einsehen.

A) Merkmalsbasierte Klassifikation

Implementieren Sie ein ML-System zur Klassifikation der Hunderassen mithilfe eines k -Nearest Neighbor Klassifikators und dessen Implementierung aus der Python-Bibliothek Scikit Learn. Bearbeiten Sie dabei die folgenden Schritte:

1. Laden der Daten und Splitting in Trainings- und Testdaten (Verhältnis 75:25).
2. Überlegen Sie sich geeignete Merkmale, die Sie aus den Bildern extrahieren können und extrahieren Sie die Merkmale.
3. Wählen Sie zur Merkmalsreduktion je eine Methoden aus dem Bereich Merkmalstransformation und Merkmalsselektion aus. Geben Sie in Ihrem Kurzreport an, aus welchem Grund Sie die jeweilige Methode gewählt haben. Vergleichen Sie am Ende die Performance Ihres ML-Systems ohne Reduktion, mit der gewählten Transformation und der gewählten Selektion.
4. Implementieren Sie mithilfe von Scikit Learn einen k -Nearest Neighbor Klassifikators zur Lösung der Klassifikationsaufgabe. Führen Sie eine Hyperparameter-Optimierung in Bezug auf k durch und stellen Sie die Erkenntnisse in Ihrem Kurzreport dar
5. Evaluieren Sie Ihr ML-System und stellen Sie die Ergebnisse in geeigneter Weise dar. Nutzen Sie zur Evaluation eine Kreuzvalidierung (3-fold).

B) Deep Learning

Implementieren Sie ein ML-System zur Klassifikation der Hunderassen mithilfe von Deep Learning. Nutzen Sie dafür Implementierungen in Keras/Tensorflow 2. Bearbeiten Sie dabei die folgenden Schritte:

1. Laden der Daten und Splitting in Trainings- und Testdaten (Verhältnis 75:25).
2. Überlegen Sie sich eine geeignete Netzwerkarchitektur für die Klassifikationsaufgabe.
3. Implementieren Sie Ihr Netzwerk in Keras/Tensorflow.

4. Wählen Sie eine geeignete Loss-Funktion und einen geeigneten Optimizer. Begründen Sie Ihre Auswahl in Ihrem Kurzreport.
5. Evaluieren Sie Ihr ML-System und stellen Sie die Ergebnisse in geeigneter Weise dar. Nutzen Sie zur Evaluation eine Kreuzvalidierung (3-fold).
6. Wählen Sie nun 5 andere Hunderassen aus. Nutzen Sie Transfer Learning zur Klassifikation dieser Rassen, d.h. laden Sie die Gewichte aus Ihrem in Teilaufgabe B5 trainierten Modell als initiale Gewichte für das neue Training. Beschreiben Sie ihr Vorgehen im Kurzreport und setzen Sie die Evaluation dieses erneuten Trainings in Relation zu den Evaluationsergebnissen ihres ursprünglichen Modells.