

Programmierabgabe 1 - Gruppe 5

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.

Aufgabe

Implementieren Sie ein ML-System zur Klassifikation der Hunderassen mithilfe einer Support Vector Machine (SVM) und deren 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 eine geeignete Methoden aus. Falls Sie mehrere Methoden vergleichen, bevor Sie sich für eine endgültige entscheiden, vergleichen Sie diese miteinander und begründen Sie Ihre Entscheidung.
4. Implementieren Sie mithilfe von Scikit Learn eine SVM zur Lösung des Klassifikationsproblems. *Hinweis: verwenden Sie keine lineare SVM, sondern eine mit Radial Basis Function Kernel, da die sich besser zur Trennung von nicht trivial trennbaren Klassen eignet.*
5. Evaluieren Sie Ihr ML-System und stellen Sie die Ergebnisse in geeigneter Weise dar. Nutzen Sie zur Evaluation eine Kreuzvalidierung (3-fold).