

Machine Learning Lab Zweitversuch

Dieses Übungsblatt dient zur Kompensation von Lab-Inhalten. Die Aufgaben 1 bis 3 widmen sich den Themen, die in den drei Lab-Blöcken bearbeitet wurden. Das Übungsblatt gibt insgesamt 100 Punkte, wobei ihr selbst wählen könnt, wo (und ob) ihr bei eurer Ausarbeitung Schwerpunkte setzen wollt.

Bei der Ausarbeitung könnt ihr euch natürlich auf unsere Mega-Tutorials und die Musterlösungen der ursprünglichen Übungsblätter stützen. Abzugeben ist jeweils eure dokumentierte Lösung als **jupyter notebook**. Bitte ladet diese als Archiv (rar/zip) in **Moodle** hoch.

Alle Aufgaben basieren auf dem **car.csv** Datenset, das wir euch in unserem **Kursrepository auf GitLab** bereitgestellt haben. Neben dem Datenset selbst findet ihr dort auch eine Detaillierte Beschreibung der Daten (**car_description.txt**).

Teil 1: Supervised Regression Learning (30+5 Punkte)

Verwendet in diesem Teil des Übungsblatts das Datenset um (mindestens) zwei Modelle zu entwickeln, um das Versicherungsrisiko (**insurance risk**) von Fahrzeugen zu schätzen. Verwendet dazu zwei unterschiedliche Regressionsansätze aus dem Repertoire von **sklearn** und dokumentiert euer vorgehen. Achtet dabei auf eine saubere Trennung von Trainings- und Testdaten. Interpretiert anschließend eure Ergebnisse/Scores.

Ihr könnt bei der Vorverarbeitung der Daten mit oder ohne **sklearn** Pipelines arbeiten. Wenn ihr die Aufgabe mit Pipelines umsetzt, bekommt ihr 5 Bonuspunkte

Teil 2: Supervised Segmentation Learning (30+5 Punkte)

Verwendet in diesem Teil des Übungsblatts das Datenset um die Zylinderzahl der Fahrzeuge zu schätzen. Behandelt die Variable **num-of-cylinders** dazu als kategoriales Merkmal. Ihr könnt dabei auf die Vorverarbeitungsschritte aus dem ersten Teil (soweit möglich) aufbauen. Entwickelt auch hier (mindestens) zwei Modelle auf Basis unterschiedlicher **sklearn** Klassifizierungsalgorithmen. Achtet auch hier auf eine saubere Trennung von Trainings- und Testdaten und interpretiert eure Ergebnisse/Scores.

Ihr könnt bei der Vorverarbeitung der Daten mit oder ohne **sklearn** Pipelines arbeiten. Wenn ihr die Aufgabe mit Pipelines umsetzt, bekommt ihr 5 Bonuspunkte

Teil 3: Supervised & Unsupervised Deep Learning (30 Punkte)

Verwendet im letzten Teil des Übungsblatts entweder eure Lösung aus Teil 1 oder aus Teil 2 und ersetzt eure ursprünglich entwickelten Modelle durch ein künstliches Neuronales Netz (Fast Forward) mit zumindest 2 hidden Layern. Verwendet hierfür **tensorflow** und **keras**. Testet die Performanz eures Netz mit Hilfe eurer Testdaten und diskutiert eure Ergebnisse. Auch hier könnt ihr, soweit möglich, auf die Vorverarbeitungsschritte eurer ersten Lösung zurückgreifen.