

Einführung in Deep LearningBlatt 1 - Abgabe am Mi, 16.10.2024 um 12:00

Wenn Sie mindestens die Hälfte der Punkte erreicht haben, gilt dieses Blatt als bestanden. Wenn Sie 8 der 10 Übungsblätter bestanden haben, haben Sie die Klausurzulassung erreicht.

Dieses Blatt können Sie allein oder in 2er- oder 3er-Teams abgeben. Achtung: Ab Blatt 2 können Sie nicht mehr alleine abgeben (Alleinabgaben werden nicht berücksichtigt).

Geben Sie Ihre Lösung als `blatt1.py` oder `blatt1.ipynb` ab.

Aufgabe 1. Implementieren Sie ein Perzeptron zur binären Klassifikation mit Python und Numpy. Ihr Code sollte eine Methode `train(x, y)` haben, die einen Numpy-Array x an Eingabedaten und einen gleich langen Numpy-Array y an Labels bekommt und eine Methode `infer(x)`, die einen Numpy-Array x an Eingabedaten bekommt und Labels dazu zurück gibt. Sie können sich die Aufgabe erleichtern, wenn Sie nur $x_i \in \mathbb{R}^2$ zulassen, also dass die Instanzen der Daten jeweils zweidimensional sind (und damit ist x eine Punktmenge in der Ebene).

Aufgabe 2. Generieren Sie geeignete synthetische Daten für eine binäre Klassifikation. Dazu dürfen Sie auch Scipy und Scikit-learn verwenden. Ihr Code sollte das in einer Methode `generate` tun. Sie haben bei der Verteilung der Daten freie Hand.

Aufgabe 3. Evaluieren Sie Ihre Implementierung des Perzeptrons systematisch anhand der synthetischen Daten und einem Train-Test-Split. Dazu dürfen Sie auch Scipy und Scikit-learn verwenden. Ihr Programm sollte das bei Aufruf von `python blatt1.py` ausführen und mit `print` den Fehler (Anteil der falsch klassifizierten) ausgeben. In einem Notebook sollte das Ergebnis nach Ausführen aller Zellen zu sehen sein.

Aufgabe 4. Evaluieren Sie Ihre Implementierung des Perzeptrons anhand des Iris-Flower-Datasets, aus dem Sie nur die Spezies 'setosa' und 'versicolor' auswählen. Dazu dürfen Sie auch Scipy und Scikit-learn verwenden. Geben Sie mit `print` den Fehler auf dem gesamten Datensatz nach Training auf dem gesamten Datensatz aus.

Wenn Sie sich das ganze visualisieren möchten, empfiehlt sich, aus dem Iris-Flower-Datensatz zwei Attribute auszuwählen. Achtung: es macht einen Unterschied, ob der ganze Datensatz (mit 4 Attributen) zum Training verwendet wird und nur die Visualisierung in 2 Dimensionen stattfindet, oder ob bereits das Training mit nur 2 Attributen durchgeführt wird. Überlegen Sie sich, warum.

Wenn Sie also alle Aufgaben erledigt haben, sollte ihr Programm (`.py`) bei Aufruf 2 Zeilen ausdrucken. Wenn Sie ein Notebook (`.ipynb`) abgeben, können Sie auch Visualisierungen mit Matplotlib einbinden.