

Machine Learning (Online)

Übung 2

Ziel dieser Übung ist die Klassifikation von Immobilien mit Hilfe einer logistischen Regression und mit einem Neuronalen Netzwerk. Auf AD finden Sie drei Datensätze, `house_train_data.csv` (zum Trainieren der Modelle), `house_test_data.csv` (zum Evaluieren resp. Testen) und `dev_data.csv` (ein kleines Datenset zum Entwickeln).

Implementieren Sie die Algorithmen jeweils selber und zeigen Sie zuerst anhand des kleinen Entwicklungsdatensatzes (`dev_data.csv`), dass

- Ihr Algorithmus konvergiert durch Zeichnen von *Cost Function* vs. *Number Iterations*.
- die Regularisierung funktioniert,
- die Klassifikation/Vorhersage funktioniert,
- die zu berechnenden Kenngrößen den Erwartungen entsprechen.

Für diese Überprüfungen erhalten Sie jeweils den ersten Teil der aufgeführten Punkte.

Aufgaben:

- 1) **(5 + 5 Punkte)** Bauen Sie mit logistischer Regression mindestens drei verschiedene sinnvolle Modelle (Anzahl Features, mit/ohne Regularisierung, etc.) zur Klassifikation der Immobilien in zwei Klassen ('waterfront': True/False). Bestimmen Sie für diese Modelle jeweils auf den Testdaten die *Confusion Matrix*, die Kenngrößen *Precision*, *Recall*, *F₁-Score* und *Accuracy*. Geben Sie eine kurze Interpretation der Resultate.

→ https://en.wikipedia.org/wiki/Confusion_matrix

→ https://en.wikipedia.org/wiki/Precision_and_recall

- 2) **(7 + 8 Punkte)** Bauen Sie mit Hilfe Neuronaler Netzwerke mindestens drei verschiedene, sinnvolle Klassifikatoren (verschiedene Anzahl Nodes im *Hidden Layer*, allenfalls mehrere *Hidden Layers*, mit/ohne Regularisierung, etc.) zur Voraussage des Gebäudezustands ('condition' ∈ {1,2,3,4,5}). Berechnen Sie jeweils mit den Testdaten die *Confusion Matrix* und die *Overall Accuracy*. Geben Sie eine kurze Interpretation der Resultate.

Überlegen Sie sich, wie Sie Kenngrößen *Precision*, *Recall*, etc. für eine *multiclass* Klassifizierung verallgemeinern würden. Siehe dazu und zu *Overall Accuracy*:

→ <https://www.youtube.com/watch?v=FAr2GmWNbT0>

Abgabe: Beschreiben Sie Ihre Lösung/Modelle in einem PDF und senden Sie dieses zusammen mit dem Code als Zip-Datei in der Form

Vorname_Name.zip

bis spätestens 30. April 2018 an manuel.galbier@fhnw.ch.