

Oefeningen Neurale Netwerken

Het doel van de onderstaande opdrachten is dat je vooral experimenteert met de neurale netwerken. Alleen zo kan je meer inzicht en intuïtie krijgen in de werking ervan. Daarom zijn er soms meerdere oplossingen mogelijk. Er zijn echter ook wel meerdere manieren waarop je dingen verkeerd kan doen...

Tips:

- Inputlagen hebben doorgaans meer neuronen nodig.
- Normaliseer of standaardiseer de input- en outputvalues (zoniet worden de activatiefuncties “verzadigd” en verloopt het leren trager).
- Elke bijkomstige laag kan “features” leren die verscholen zitten in de laag ervoor.
- Begin klein zowel qua lagen als qua aantal neuronen en breid uit wanneer je geen resultaten behaalt. Een vuistregel laat het aantal neuronen in opeenvolgende lagen afnemen volgens een macht van 2, bijv. 64, 32, 16, ...

Vraag 1 – The Simpsons Revisited

We willen het geslacht van een Simpson kunnen voorspellen m.b.v. een neurale netwerk (cfr. hoofdstuk Beslissingsbomen uit Data Science 1). Deze dataset is erg klein en daarom kunnen we geen validatie-set of test-set construeren.

- a) Lees de dataset ('The Simpsons original.csv') van de Simpsons in.
- b) Wat voor soort voorspelling trachten we hier te doen? Regressie of classificatie?
- c) Bepaal welke kolommen we niet kunnen gebruiken bij deze voorspelling?
- d) Maak nu een input training data set `x_simpsons` en normaliseer deze (min-max normalisatie). Doe hetzelfde voor de outputkolom (`y4_simpsons`).
- e) Maak een neurale netwerk met de correcte inputs en outputs. Maak een plot van het ANN.
- f) Gebruik je neurale netwerk om het geslacht van je `x_simpsons` te voorspellen. Komen ze overeen met de targets? Voorspel ook het geslacht van de onbekende Simpson (Comic guy: `hairlength=8`, `weight=500`, `age=38`).. Komt dit overeen met de voorspelling die je gemaakt werd bij beslissingsbomen (zie Data Science 1)?

Vraag 2 – Forecast demo Revisited

We willen de opbrengsten van een bedrijf per kwartaal van de laatste 5 jaar kunnen voorspellen (cfr. hoofdstuk Forecasting uit Data Science 1). Deze dataset is erg klein en daarom kunnen we geen validatie-set of test-set construeren.

- a) Lees de dataset van de Forecasting in (forecastdemo.csv)
- b) Wat voor soort voorspelling trachten we hier te doen? Regressie of classificatie?
- c) Bepaal welke kolommen we niet kunnen gebruiken bij deze voorspelling?
- d) Maak nu een input training data set `x_forecast` en een outputkolom (`y_forecast`). Normaliseer waar nodig.
- e) Maak een neuraal netwerk met de correcte inputs en outputs. Maak een plot van de ANN.
- f) Gebruik je neuraal netwerk om de opbrengsten te voorspellen die horen bij `x_forecast`. Komen ze overeen met de werkelijke opbrengsten? Maak één plot van de originele gegevens en de voorspelde gegevens

Vraag 3 – Iris

- a) Laad de dataset iris (iris.csv) op en bestudeer de inhoud.
- b) Wat voor soort voorspelling trachten we hier te doen? Regressie of classificatie?
- c) Bepaal welke kolommen we niet kunnen gebruiken bij deze voorspelling?
- d) Maak nu een input data set `x_iris` en normaliseer deze. Probeer een van de normalisatietechnieken uit die je nog niet hebt gebruikt in een oefening of opdracht.
- e) Creëer een trainig data set en een test data set (`x_train_iris` en `x_test_iris`) die respectievelijk bestaan uit 80% en 20% (willekeurig gekozen) van de input data set `x_iris`.
- f) Maak een neuraal netwerk met de correcte inputs en outputs.
- g) Evalueer je neuraal netwerk op je test data set (`x_test_iris`). Welke nauwkeurigheid (accuracy) behaal je?