

Titel

Roel van Warmerdam: 4300556, Lukas Donkers: ?

June 24, 2016

1 inleiding

Een inleiding waarin de probleemstelling wordt uiteengezet. Wat is de analysevraag, en wat is de beoogde toepassing van de modellen? Hierbij mag je natuurlijk delen van de opdrachtbeschrijving hergebruiken, maar probeer zoveel mogelijk zelf te formuleren.

2 beschrijving

Een beschrijving van de beschikbare data, inclusief eenvoudige beschrijvende statistieken. Meld ook zaken die je opvallen bij het "inspecteren" van de data, bijvoorbeeld zaken die op "vervuiling" kunnen wijzen.

3 voorbewerkingen

Beschrijf eventuele voorbewerkingen die je op de tekst hebt uitgevoerd, bijvoorbeeld het verwijderen van leestekens en stopwoorden, stemming, etc.

4 analyse

Beschrijf nauwkeurig welke features je hebt bedacht, en waarom je denkt dat die features voorspellende waarde zouden kunnen hebben. Toon vervolgens de output van het lineaire regressiemodel met die features, getraind op de trainingset. Zijn de tekens van de coëfficiënten wat je had verwacht? Welke coëfficiënten zijn significant bij $\alpha=0.05$? Doe vervolgens een local search met

stepAIC, waarbij ook interacties tussen 2 features mogen worden toegevoegd. Bereken van beide modellen de root mean square error op de testset. Doe een soortgelijke analyse met ordinale logistische regressie en multinomiale logistische regressie in plaats van lineaire regressie, en rapporteer de resultaten.