

Leren: Programmeeropdracht 4

Week 5 - Semester 1 2014/15

Mag in teams van 2 gemaakt worden
Deadline 1 december 2014 13.00 uur

1. (2 punt) *K-nearest neighbour*

- (a) Implementeer het k-nearest neighbour algoritme. Ga uit van de Euclidische afstand. Pas het toe op de trainingsdata digits123-1 en test het op digits123-2. Bereken de proportie goede classificaties.
- (b) Maak een variant die neighbours weegt met de afstand tot het testvoorbeeld. Ga na of dit beter werkt dan de basisversie.
- (c) Maak een variant die de variabelen weegt met hun voorspellende waarde. Gebruik hiervoor de accuracy.

2. (4 punten) *Vergelijken van algoritmen*

Ga uit van de written digits data digits123-1 en digits123-2 uit de vorige deelopdracht.

- (a) Welke algoritmen die we tot nu toe gebruikt hebben zijn geschikt voor dit probleem?
- (b) Ga uit van de accuracy als evaluatiemaat. Hoe goed een algoritme werkt meten we af aan de accuracy op de digits123-2. Implementeer een programma dat de accuracy van de voorspellingen van een hypothese uitrekent na het leren, of gebruik dat uit de vorige opdracht.
- (c) Probeer een zo goed mogelijke cijferherkenner te vinden voor dit probleem *in zo min mogelijk stappen*. Een stap is het runnen van 1 leeralgoritme op (een deel van) de data. Het kan gaan om een ander algoritme, maar ook om andere (meta)parameters of het weglaten of toevoegen van variabelen. Inleveren: verslag van stappen die je hebt gezet, steeds de reden waarom je deze stap hebt gedaan en de resulterende accuracy op digits123-2. (NB: als je in een stap alleen iets probeert, zet dat er dan bij).