

# Vertaling Eindtentamen Machine Learning

## Normale Groep, Normale versie

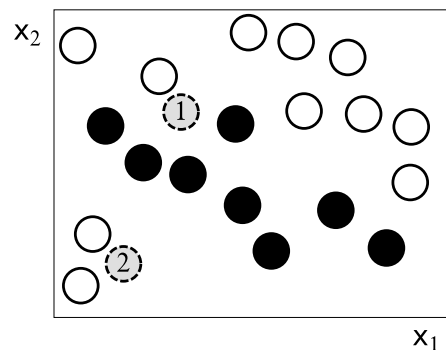
20 december, 2007

18.30 – 21.15

N.B. Deze vertaling dient alleen ter ondersteuning bij het begrijpen van de Engelse versie; Er kunnen geen rechten aan worden ontleend. Bij de beoordeling wordt uitsluitend uitgegaan van de Engelse versie. Eventuele fouten in deze vertaling zijn geen reden voor een soepelere beoordeling.

Schrijf je tentamenversie op je antwoordenblad! Het gebruik van een rekenmachine is toegestaan. Het tentamen wordt als volgt nagekeken: Je begint met 1 punt, en met elk van de 12 deelvragen kun je 3/4 punt verdienen. Voor antwoorden die ten dele goed zijn kan een deel van de punten worden toegekend. Succes!

- (a) Figuur 1 toont classificatiedata met twee klassen: Zwart en Wit. De twee gevallen met gestippelde lijnen, die genummerd zijn als 1 en 2, zijn nog niet geïdentificeerd. Welke klassenlabels zouden hieraan worden toegekend door  $k$ -nearest neighbour met  $k = 1$ ,  $k = 3$  en  $k = 5$ ?



Figuur 1: Een classificatiedataset

- (b) In een andere dataset, gegeven in Tabel 1, heeft het attribuut (feature)  $x_1$  drie mogelijke waarden: Zwart, Wit en Bruin. Het attribuut  $x_2$  kan iedere geheeltallige waarde aannemen, en je mag aannemen dat het verschil tussen twee van zijn waarden een zinnige betekenis heeft. Wat zou een geschikte manier zijn om deze attributen te representeren voor het  $k$ -nearest neighbour algoritme (aannemend dat het Euclidische afstand gebruikt tussen feature vectoren)?

Tabel 1: Een dataset

$x_1$ PaardKleur	$x_2$ AantalVijandenVerlagen	$y$ GoedOfKwaad
Zwart	1	Goed
Zwart	36	Kwaad
Wit	0	Goed
Bruin	0	Goed

- (c) Misschien zou het beter zijn om AantalVijandenVerlagen in do-zijnen uit te drukken. Dus stel dat we  $x_2$  van Tabel 1 in eenheden van twaalf zouden uitdrukken. Bijvoorbeeld,  $1 \rightarrow 1/12$  en  $36 \rightarrow 3$ . Welk effect zou dit hebben op het  $k$ -nearest neighbour algoritme?
  - (d) Stel je nog weer een andere classificatietask voor, waarin er twee attributen zijn,  $x_1$  en  $x_2$ , die beide de waarden  $1, 2, \dots, 100$  kunnen aannemen, en waarin de mogelijke klassen weer Zwart en Wit zijn. Stel dat de target functie het klassenlabel  $y$  toekent als op een schaakbord, waarbij  $x_1$  de rijen van het bord indexeert en  $x_2$  de kolommen:  $y$  is Zwart als  $x_1 + x_2$  even is en  $y$  is Wit als  $x_1 + x_2$  oneven is. Zou het moeilijk of makkelijk zijn (in termen van de hoeveelheid benodigde trainingsdata) for 1-nearest neighbour om deze target functie te leren? (Motiveer je antwoord.)
- (a) Stel dat we perceptron gewichten willen leren van de volgende vijf voorbeelden:

$$D = \begin{array}{c|ccccc} y & -1 & -1 & -1 & 1 & -1 \\ \hline x_1 & -1 & 1 & 0 & 1 & -1 \\ \hline x_2 & 1 & -1 & 0 & 1 & -1 \end{array}$$

Laat zien dat er gewichten bestaan zodat het perceptron alle voorbeelden correct classificeert.

- (b) Geef een voorbeeld van een dataset met tenminste vier voorbeelden, waarvan een perceptron er altijd tenminste één verkeerd zou classificeren.
3. In een bepaalde uitbreiding van het basialgoritme voor gradient descent laat men de learning rate langzaam afnemen terwijl het algoritme draait. Waarom zou men dat doen?
4. Hoe zou naive Bayes, gegeven de trainingsdata in Tabel 2, een nieuwe feature vector classificeren waarvan beide componenten de waarde True hadden?

Tabel 2: Enige Boolean-waardige data

$x_1$	$x_2$	$y$
False	False	False
False	True	True
True	False	True

5. Stel dat we willen voorspellen hoe de volgende binaire reeks verder gaat:

$$D = \begin{array}{|c|c|c|c|c|c|c|c|c|} \hline y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & \\ \hline 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & \\ \hline \end{array}.$$

Gegeven is een hypotheseruimte met twee deterministische hypothesen,  $\mathcal{H} = \{h_1, h_2\}$ , die als volgt voorspellen:

$$h_1 : y_n = 1,$$

$$h_2 : \begin{cases} y_n = 0 & \text{voor } n \leq 3, \\ y_n = 1 & \text{voor } n > 3. \end{cases}$$

- (a) Stel dat we denken dat de kans om een meetfout te observeren  $1/10$  is voor iedere uitkomst  $y_n$  in de reeks. Beschrijf het model waarin deze kennis is verwerkt.
- (b) Welke hypothese uit dat model zou geselecteerd worden door maximum likelihood parameter estimation gebaseerd op data  $D$ ? (Voeg voldoende berekeningen bij om je antwoord te onderbouwen.)
- (c) Welke hypothese uit dat model zou geselecteerd worden door Bayesiaanse MAP estimation als we prior kans  $1/100$  zouden geven aan de hypothese die maximum likelihood selecteert en  $99/100$  aan de andere hypothese in het model? (Voeg voldoende berekeningen bij om je antwoord te onderbouwen.)

6. Stel we hebben een Engelse tekst van  $n$  woorden en we willen two-part MDL gebruiken om te kiezen tussen de drie context-vrije grammatica's (context-free grammars, CFGs) van het college:

- de 'promiscuous' grammatica, die iedere tekst van iedere lengte accepteert;
- de ad hoc grammatica, die alleen de trainingstekst accepteert;
- de 'correcte' ('right') grammatica, die een goede CFG-benadering vormt van de werkelijke grammatica van het Engels.

Waarom prefereert two-part MDL de 'correcte' grammatica boven de andere twee grammatica's? Vermeld in je antwoord of  $L(H)$  en  $L(D | H)$  klein of groot zijn voor de grammatica's ten opzichte van de grootte van de ongecomprimeerde tekst (aannemend dat  $n$  zeer groot is).