

Herkansingstentamen Machine Learning

14 Februari, 2008

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.

Geef voldoende onderbouwing voor je antwoorden. 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!

Tabel 1: Classificatiedata

x_1	x_2	y
0	2	+1
2	2	-1
1	2	-1
0	0	+1

1. Stel we hebben een classificatietask waarin het klassenlabel y de waarden -1 en $+1$ kan aannemen, en waarin er twee features, x_1 en x_2 , zijn die beide als mogelijke waarden 0 , 1 en 2 hebben. Laat $\mathcal{H} = \{h_1, h_2, h_3\}$ een hypotheseruimte zijn voor dit probleem die de volgende drie hypo-

thesen bevat¹:

$$h_1(\mathbf{x}) = \begin{cases} +1 & \text{als } x_1 \cdot x_2 = 0, \\ -1 & \text{in alle andere gevallen.} \end{cases}$$

$$h_2(\mathbf{x}) = \begin{cases} +1 & \text{als } x_1 \neq x_2, \\ -1 & \text{in alle andere gevallen.} \end{cases}$$

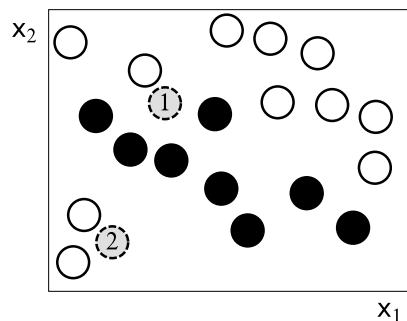
$$h_3(\mathbf{x}) = \begin{cases} +1 & \text{als } x_1 = 0, \\ -1 & \text{in alle andere gevallen.} \end{cases}$$

- (a) Geef een beslisboom (decision tree) die hetzelfde classificeert als h_1 .
 - (b) Hoe zou het LIST-THEN-ELIMINATE algoritme, gebruikmakend van hypotheseruimte \mathcal{H} , het nieuwe geval $\mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ classificeren gebaseerd op de data in Tabel 1? (Vergeet niet je antwoord te motiveren.)
 - (c) Geef een voorbeeld van een hypothese voor dit classificatieprobleem die consistent is met de data in Tabel 1, maar geen element is van \mathcal{H} .
 - (d) Welke van de hypothesen in \mathcal{H} kan geïmplementeerd worden door een perceptron door geschikte gewichten te kiezen?
 - (e) Geef een voorbeeld van een prefixcode, zoals die gebruikt worden in minimum description length leren, om de elementen van \mathcal{H} te coderen.
2. Hoe zou naive Bayes het nieuwe geval $\mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ classificeren gegeven de trainingsdata in Tabel 1? (Geef voldoende berekeningen om je antwoord te onderbouwen.)

Tijdens het tentamen werd er een fout gevonden in deze opgave: Naive Bayes kan het gegeven voorbeeld $\mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ niet classificeren, omdat in Tabel 1 de feature x_2 voor geen enkele van de klassen de waarde 1 krijgt. Daarom werd de vraag veranderd: In the gecorrigeerde versie werd gevraagd hoe naive Bayes het nieuwe geval $\mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ zou classificeren.

¹De notatie $x_1 \cdot x_2$ staat voor het product van x_1 en x_2 .

3. Stel we krijgen een nieuwe dataset D en draaien er het ID3 algoritme op.
- Als we de boom die ID3 selecteert zouden gebruiken om nieuwe data te classificeren, kunnen we dan ongeveer hetzelfde foutpercentage verwachten als op data D ? (Motiveer je antwoord.)
 - Stel dat we na het draaien van ID3 reduced-error pruning toepassen, waarbij we echter dezelfde data D gebruiken om te bepalen welke knopen (nodes) in de beslisboom we prunen. Wat zou het effect van prunen in dit geval zijn?
4. (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 geclassificeerd. Welke klassenlabels zouden hieraan worden toegekend door k -nearest neighbour met $k = 1$, $k = 3$ en $k = 5$?



Figuur 1: A classification data set

- Stel dat we alle featurewaarden met hetzelfde getal vermenigvuldigen, wat zou dan het effect zijn op het k -nearest neighbour algoritme (aannemend dat het Euclidische afstand tussen feature vectors gebruikt)? (Motiveer je antwoord.)
- 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 volgens

de volgende regel: y is Zwart als $x_1 + x_2 > 100$ en y is Wit in alle andere gevallen. Zou het moeilijk of makkelijk zijn (in termen van de hoeveelheid benodigde trainingsdata) voor 5-nearest neighbour om een goede benadering van deze target functie te leren? Bespreek in je antwoord hoe goed de geleerde benadering zou zijn: Zijn er bepaalde inputs waarop het waarschijnlijker is dat de benadering een fout maakt? (Motiveer je antwoorden.)

5. Stel we hebben de volgende data, die bestaat uit a 's en b 's:

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

We krijgen een model met twee probabilistische hypothesen, $\mathcal{M} = \{P_\theta \mid \theta \in \{1, 2\}\}$, die de volgende voorspellingen doen:

$$\begin{array}{ll} P_1(y_n = a) = 0.3 & P_2(y_n = a) = 0.8 \\ P_1(y_n = b) = 0.7 & P_2(y_n = b) = 0.2 \end{array}$$

Verzin een prior distributie op θ zodanig dat voor data D Bayesiaans MAP een andere hypothese selecteert dan maximum likelihood parameter estimation. (Geef berekeningen die laten zien welke hypothesen geselecteerd worden door maximum likelihood en MAP.)