

Vertaling Eindtentamen Machine Learning

Normale Groep,

Herkansing-van-het-deeltentamen 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!

- In Hoofdstuk 2 van Mitchell staat de volgende hypotheseruimte beschreven voor het EnjoySport concept learning voorbeeld:

$$\mathcal{H} = \{ \langle ?, ?, ?, ?, ? \rangle, \langle \text{Sunny}, ?, ?, ?, ? \rangle, \langle \text{Cloudy}, ?, ?, ?, ? \rangle, \dots, \langle \emptyset, \emptyset, \emptyset, \emptyset, \emptyset \rangle \}$$

Ter herinnering: Iedere hypothese bestaat uit een lijst van zes beperkingen (constraints), een voor ieder attribuut. Als een feature vector voldoet aan alle beperkingen, dan wordt hij geassocieerd als Yes (onderdeel van het concept), anders als No (geen onderdeel van het concept). Het eerste attribuut heeft drie mogelijke waarden, de andere attributen twee. Iedere beperking kan de volgende waarden aannemen:

Waarde van Beperking	Omschrijving
?	Het betreffende attribuut accepteert iedere mogelijke waarde.
\emptyset	Geen enkele waarde wordt geaccepteerd.
<i>Warm</i>	Specifieke waarde vereist voor het attribuut (bijv. Warm)

- Bepaal voor ieder van de volgende hypothesen of ze in deze hypotheseruimte \mathcal{H} zitten. Zo ja, druk de hypothese dan uit in Mitchells notatie als een lijst van zes beperkingen. Zo nee, leg uit waarom de hypothese niet in deze vorm kan worden uitgedrukt.

$$h_1(\mathbf{x}) = \begin{cases} \text{Yes} & \text{als } x_2 = \text{'Warm'} \text{ en } x_3 = \text{'Normal'}, \\ \text{Yes} & \text{als } x_2 = \text{'Cold'} \text{ en } x_3 = \text{'Normal'}, \\ \text{No} & \text{in alle andere gevallen.} \end{cases}$$

$$h_2(\mathbf{x}) = \begin{cases} \text{Yes} & \text{als } x_1 = \text{'Sunny'} \text{ en } x_2 = \text{'Cold'}, \\ \text{Yes} & \text{als } x_1 = \text{'Sunny'} \text{ en } x_2 = \text{'Warm'} \text{ en } x_3 = \text{'High'}, \\ \text{No} & \text{in alle andere gevallen.} \end{cases}$$

- Geef een voorbeeld van een hypothese die niet in \mathcal{H} zit. (Een uitleg is niet vereist.)

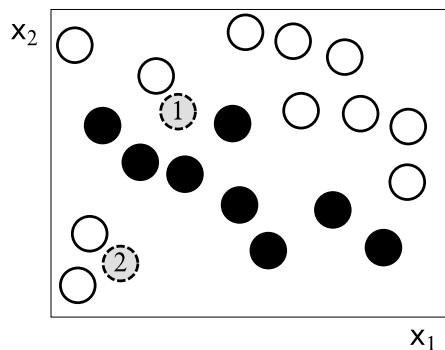
- Hoe bepaalt ID3 welke attributen het dichtbij de root van de boom (tree) plaatst?

- Wanneer een 14de-graads polynoom op 15 datapunten wordt gepast (gefit) met behulp van least squares regressie, zal er overfitting optreden. Beredeneer voor ieder van de volgende twee gevallen of ze zouden helpen tegen overfitting:

- Stel dat we in plaats van de som van de kwadraten van de fouten, de som van de absolute waarden van de fouten zouden minimaliseren.
- Stel dat we in plaats van 15 datapunten 100 000 datapunten hadden.

- (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 geassocieerd. Welke klassenlabels zouden hieraan worden toegekend door k -nearest neighbour met $k = 1$, $k = 3$ en $k = 5$?

- 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



Figuur 1: Een classificatiedataset

(aannemend dat het Euclidische afstand gebruikt tussen feature vectoren)?

Tabel 1: Een dataset

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

5. Geef een voorbeeld van een dataset met tenminste vier voorbeelden, waarvan een perceptron er altijd tenminste één verkeerd zou classificeren. (N.B. Je antwoord moet dus de volgende vorm hebben:

$$\begin{pmatrix} y_1 \\ \mathbf{x}_1 \end{pmatrix}, \begin{pmatrix} y_2 \\ \mathbf{x}_2 \end{pmatrix}, \begin{pmatrix} y_3 \\ \mathbf{x}_3 \end{pmatrix}, \begin{pmatrix} y_4 \\ \mathbf{x}_4 \end{pmatrix}, \dots,$$

met specifieke getallen ingevuld voor iedere y en de componenten van iedere \mathbf{x} .)

6. Hoe zou naive Bayes, gegeven de trainingsdata in Tabel 2, een nieuwe feature vector classificeren waarvan beide componenten de waarde True hadden?

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

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

Gegeven is een model met twee probabilistische hypothesen, $\mathcal{M} = \{P_1, P_2\}$, die als volgt voorspellen:

$$P_1(y_n = 1) = 0.9$$

$$P_2(y_n = 1) = 0.1$$

- (a) 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.)
- (b) 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.)

8. 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).

Tabel 2: Enige Boolean-waardige data

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