

Toets Parel 100 — Intelligente Interactie

4 oktober 2013

Dit tentamen bestaat uit 7 vragen. Het cijfer voor deze toets is het aantal punten gedeeld door 100.

- 1 (15 punten) Twee personen wordt gevraagd een verzameling plaatjes van meubilair te annoteren met de labels "bureau", "tafel", "stoel", en "kast". Hun resultaten op de eerste 25 plaatjes zijn als volgt:

Persoon A: k, t, k, b, b, b, k, b, t, b, s, s, b, t, s, b, t, b, t, t, t, b, s, s, t

Persoon B: k, t, k, b, b, b, k, b, t, b, s, s, t, t, s, b, b, b, t, t, b, b, s, s, t

- (a) Stel een confusion matrix op voor deze twee annotaties, met de rijen en kolommen in de volgorde "bureau", "tafel", "stoel", "kast".
(b) Een confusion matrix kan je gebruiken om problemen te vinden in je annotatietaak. Welk probleem valt op in deze matrix, en waar zou dat door veroorzaakt kunnen worden?

- 2 (10 punten) Persoon C heeft ook dezelfde taak uitgevoerd. De resulterende confusion matrix tussen persoon A en persoon C ziet er als volgt uit, waarbij de rijen en kolommen weer in de volgorde "bureau", "tafel", "stoel", "kast".

```
2 3 2 0
2 1 1 3
3 0 1 0
3 3 1 0
```

- (a) Wat zegt dit over de annotaties van persoon C?

- 3 (15 punten) Deze vraag gaat over een aangepast versie van het brouwerijvoorbeeld zoals beschreven in de Machine Learning reader. Stel de de brouwerij heeft nu de volgende dataset waarbij de laatste kolom de assessment voorstelt, dus de classificatie van een locatie.

| Ex. | U | H | I | T | S | A |
|-----|---|---|---|---|---|---|
| 1 | Y | M | N | P | M | Y |
| 2 | N | S | N | P | L | N |
| 3 | Y | M | N | A | M | Y |
| 4 | N | M | N | P | S | N |
| 5 | N | M | Y | P | M | Y |
| 6 | Y | N | N | A | S | N |
| 7 | N | N | N | A | S | Y |
| 8 | N | S | N | A | M | N |
| 9 | N | L | Y | P | L | Y |
| 10 | N | M | N | P | S | N |

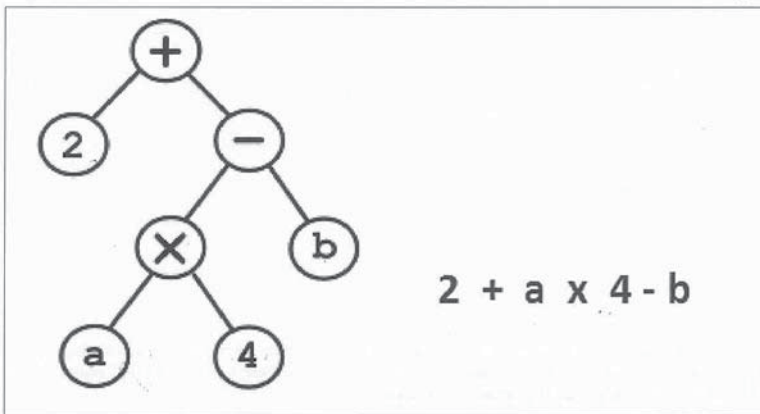
- (a) Wat is de information gain van feature T ?
(b) Stel dat T de feature is die de hoogste information gain heeft. Dus dan komt T in top van de beslisboom. Wat is de dataset die gebruikt wordt voor het bepalen van het vervolg van de beslisboom onder de tak $T = A$?
(c) Wat is de information gain voor de feature H voor de data behorende bij de tak $T = A$?

4 (15 punten) Beschouw de volgende confusion matrix behorende bij een classifier:

| | | Predicted class | |
|--------------|-------|-----------------|-------|
| | | C_1 | C_2 |
| Actual Class | C_1 | 120 | 35 |
| | C_2 | 15 | 140 |

- Wat is de accuracy van deze classifier?
- Wat is recall voor klasse C_1 ?
- Wat is de precision voor klasse C_2 ?

5 (15 punten) Onderstaande figuur toont twee presentaties van één en dezelfde aritmetische expressie.



Tot welk data type behoort, volgens de taxonomie van Shneiderman, een aritmetische expressie?

6 (10 punten) Welke van de volgende beweringen zijn waar?

- In een Network data type kunnen data items meerdere ouders (parents) hebben.
- In een Tree data type kunnen data items hoogstens één ouder (parent) hebben.
- In een Tree data is "top-down" (boven-onder) een zinvol begrip.
- In een Network data type is "top-down" een zinvol begrip.

7 (20 punten)

- Noem een mogelijke stakeholder (gebruiker) van het te ontwerpen Twitter Dashboard van de batavierenrace. (Dit mag, maar hoeft niet de functie van de stakeholder te zijn die je met je groep hebt benaderd.)
- Noem drie functies die je voor deze gebruiker op het dashboard wilt aanbieden.