

Tentamen - TA2060, Data Analysis & Geostatistics

CT 4.02, 12 augustus 2013, 14.00 - 17.00

Dit tentamen bestaat uit 45 vragen.

Je begint met 10 punten. Elk goed antwoord is 2 punten waard.

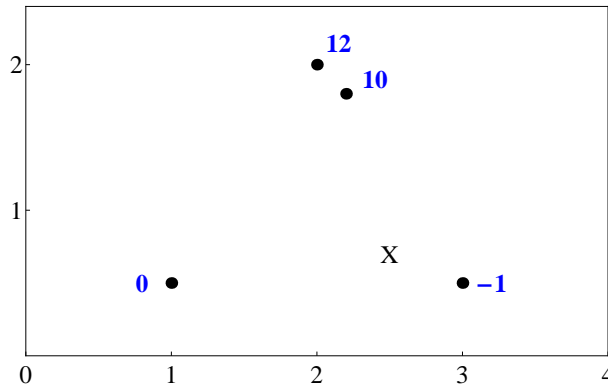
Licht altijd je antwoord toe.

Technische Engelse termen staan schuingedrukt.

Het gebruik van een rekenmachine is toegestaan.

SUCCES!

A. Deterministische interpolatie



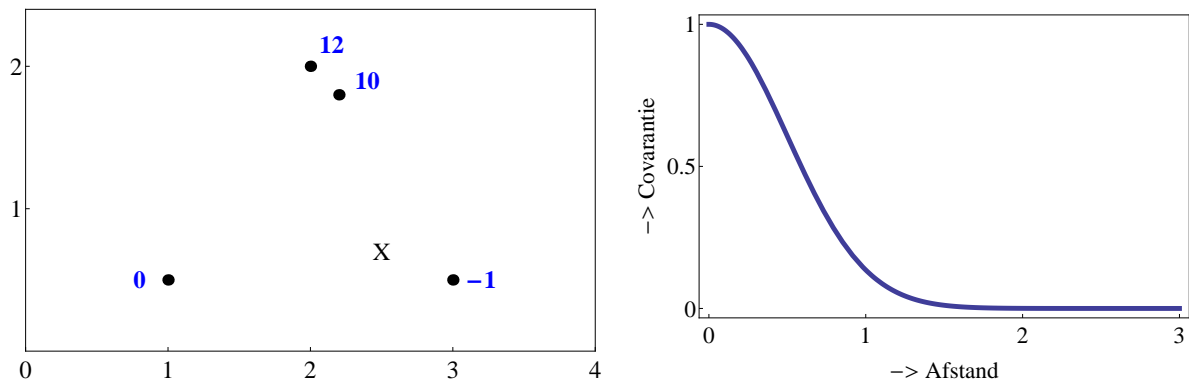
Figuur 1: Vier hoogte waarnemingen: $h_1 = 0$, $h_2 = 12$, $h_3 = 10$ en $h_4 = -1$.

In Figuur 1 staan vier hoogte waarnemingen. Henk wil de hoogte schatten op de locatie met het kruisje.

Opmerking: Bij het beantwoorden van Vragen 4 t/m 8 is het niet nodig formules te gebruiken.

1. Wat is het convex omhulsel van de waarnemingslocaties?
2. De waarnemingslocaties kunnen op twee manieren getrianguleerd worden. Welke twee manieren zijn dat?
3. Welke van deze twee manieren is een Delaunay triangulatie? Waarom? Wat is het voordeel van Delaunay triangulaties boven andere triangulaties?
4. Wat is ongeveer de hoogte bij het kruisje volgens *triangular interpolation*?
5. Welk gewicht krijgt elke waarneming in dit geval ongeveer?
6. Wat is ongeveer de hoogte bij het kruisje volgens *inverse distance interpolation* met *power 2*?
7. En welk gewicht krijgt elke waarneming ongeveer in dit geval?
8. Krijgt waarneming h_4 relatief meer of minder gewicht bij *inverse distance interpolation* met *power 1*? Waarom?
9. Teken het Voronoi Diagram van de vier waarnemingslocaties.
10. Wat is de hoogte bij het kruisje volgens *nearest neighbor interpolation*?

B. Stochastische interpolatie



Figuur 2: **Links.** Vier waarnemingen: $h_1 = 0$, $h_2 = 12$, $h_3 = 10$ en $h_4 = -1$. **Rechts.** Covariantie functie $f(x) = e^{-2x^2}$.

Ingrid gebruikt Kriging om de hoogte te schatten op de lokatie van het kruisje. Daarvoor gebruikt ze een covariantie functie $f(x)$, zie Figuur 2.

11. Welke waarnemingen zijn gecorreleerd? Waarom?
12. Wat is, in het algemeen, het verschil tussen covariantie en correlatie?
13. Wat is (ongeveer) de covariantie tussen waarnemingen h_1 en h_2 ? En wat is (ongeveer) de covariantie tussen waarnemingen h_2 en h_3 ?
14. Wat is (ongeveer) het *Kriging the mean* gemiddelde van de vier hoogte waarnemingen?

Ingrid gebruikt *Ordinary Kriging* (OK) om de hoogte te schatten bij het kruisje.

15. Wat kan er mis gaan met het oplossen van de Ordinary Kriging vergelijkingen als de covariantie functie niet positief definitief is?
16. Wat is de *range* en *sill* van de covariantie functie $f(x)$?
17. Stel de OK *redundancy matrix* op van de vier hoogte waarnemingen.
18. Wat is de *proximity vector* voor een OK schatting bij het kruisje?
19. Wat is de OK schatting van de hoogte op lokaties h_1 en h_3 ?
20. Karakteriseer de OK variantie: waar is deze over het algemeen hoog en waar laag?

C. Classificatie

Henk en Ingrid willen een combinatie van Landsat data en hoogtedata gebruiken om landgebruik in kaart te brengen. Ze hebben Landsat 7 data bestaande uit 8 banden in WGS84, een elliptisch coördinatensysteem, en een 25m raster hoogtebestand in RD, een Cartesisch coördinatensysteem.

21. Wat onderscheidt de verschillende Landsat banden?
22. Wat is een *false colour image*?
23. Wat zal ongeveer de resolutie van de Landsat data zijn?
24. Hoogtedata wordt onder andere ingewonnen met laser ranging vanuit een vliegtuig. Welke informatie over het vliegtuig is nog meer nodig om de 3D coördinaten van een punt op de grond te bepalen m.b.v. laser ranging?
25. Wat is het verschil tussen een elliptisch en een Cartesisch coördinatensysteem?
26. Waar moet je mee rekening houden als je de afstand bepaalt tussen twee punten gegeven in WGS84?
27. Waarom is het efficiënter om 10000 locaties in GIS raster formaat op te slaan dan in GIS vector formaat?

Henk en Ingrid willen behalve de hoogte ook de ruwheid van het terrein gebruiken bij de karakterisatie van het landgebruik.

28. Schets een methode om voor elke rastercel de ruwheid te bepalen aan de hand van de hoogtes in een 3×3 omgeving.
29. Schets ook een methode om de kwaliteit van de ruwheidsschatting te bepalen.

De opgave gaat verder op de volgende bladzijde

Henk en Ingrid gebruiken QGIS om de Landsat data naar RD te converteren. Als resultaat verkrijgen ze een rasterbestand met voor elke cel 8 Landsat bandwaarden en bovendien 2 geometrische waarden, hoogte en ruwheid. Henk en Ingrid willen 4 klassen onderscheiden: water, bos, landbouw en stad.

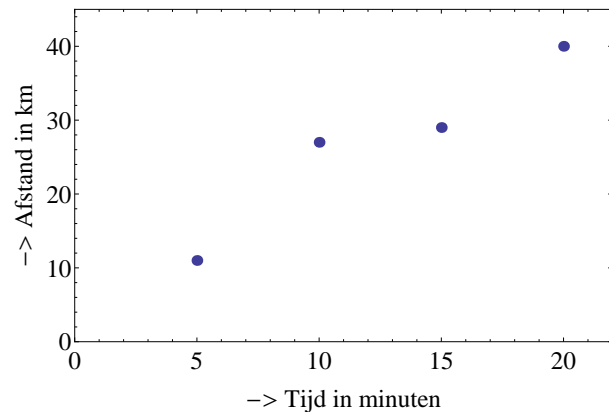
29. Waarom is Landsat data relevant voor landclassificatie? En waarom is de hoogte en de ruwheid relevant?
30. Wat is in dit geval de dimensie van de *feature space*?
31. Henk en Ingrid besluiten K-means clustering te gebruiken om de *feature space* in 4 clusters te verdelen. Leg uit hoe deze methode werkt.
32. Corresponderen de resulterende clusters noodzakelijkerwijs met de 4 te onderscheiden klassen? Waarom?

Henk en Ingrid willen hun resultaten valideren. Ze bezoeken 500 locaties in het gebied en bepalen ter plekke het landgebruik. Dit is de validatie data.

33. Leg uit hoe Henk en Ingrid de validatie data kunnen gebruiken om hun resultaten te valideren met behulp van een *confusion matrix*.
34. Geef een voorbeeld van een commissie fout.
35. 10 waarden per pixel locatie is wat onoverzichtelijk. Hoe kan je correlatie gebruiken om uit deze 10 banden de 4 banden te bepalen die het meest relevant zijn voor landclassificatie?

D. Rijsnelheid

tijd in minuten	5'	10'	15'	20'
afstand in km.	11	27	29	40



Figuur 3: Gemeten GPS afstanden.

Mark rijdt met zijn auto over een rechte weg. Elke 5 minuten bepaalt Mark de afgelegde afstand aan de hand van GPS metingen. De resultaten staan in Figuur 3. Mark wil graag zijn gemiddelde snelheid bepalen.

36. Na 10 minuten passeert Mark een gebied met veel grote kantoren. Beïnvloedt dat de GPS metingen? Waarom?

Mark besluit eerst zijn gemiddelde snelheid te bepalen door een lijn door zijn waarnemingen te schatten met de kleinste kwadraten methode.

37. Waarom gaat in dit geval de lijn door het punt $(0, 0)$?
38. Wat is de algemene vergelijking van de lijn? Hoeveel parameters moet Mark schatten?
39. Wat is de vector van waarnemingen y ?
40. Wat is de model matrix A ?
41. Geef de formule om de parameter vector \hat{x} te schatten uit de vector van waarnemingen y en de model matrix A .
42. Bereken Marks gemiddelde snelheid met de kleinste kwadraten methode.
43. Op welke afstand zou Mark zich op de vier meetmomenten bevinden volgend \hat{x} ?
44. Welke waarneming heeft het grootse residu?

Mark vertrouwt de tweede meting niet. Hij besluit daarom nogmaals zijn stijgsnelheid te schatten, maar nu met RANSAC.

45. Leg (zonder berekening) uit hoe Mark zijn gemiddelde snelheid kan bepalen met RANSAC.