

Interpolare - Metoda ponderării inverse la distanță

I. Introducere

Ponderarea inversă la distanță este o metodă deterministă pentru interpolarea multivariată, cu un set cunoscut de puncte dispuse aleatoriu (împrăștiate). Valorile alocate punctelor necunoscute sunt calculate cu o medie ponderată a valorilor disponibile în punctele cunoscute. Numele acestei metode este dat de media ponderată aplicată, deoarece recurge la inversul distanței față de fiecare punct cunoscut ("cantitate de apropiere") la atribuirea greutăților.

II. Definiția problemei

Rezultatul așteptat este o alocare discretă a funcției cunoscute u , într-o regiune de studiu:

$$u(x) : x \rightarrow \mathbb{R}, \quad x \in \mathcal{D} \subset \mathbb{R}^n, \quad \text{unde } \mathcal{D} \text{ este regiunea de studiu.}$$

Setul de date cunoscute, notat cu \mathcal{N} poate fi descris ca o lista de tuple:

$$[(x_1, u_1), (x_2, u_2), \dots, (x_N, u_N)].$$

Funcția trebuie să fie "constantă" (continuă și diferențiată), pentru a fi exactă: $(u(x_i) = u_i)$ și pentru a răspunde așteptărilor intuitive ale utilizatorului cu privire la fenomenul cercetat. Mai mult, funcția ar trebui să fie potrivită pentru o aplicație computerizată la un cost rezonabil (în prezent, o implementare de bază va folosi probabil resurse paralele).

III. Cum funcționează interpolarea ponderată la distanță inversă?

Interpolarea ponderată la distanță inversă, face în mod explicit presupunerea că lucrurile care se apropie unele de altele sunt mai asemănătoare decât cele care sunt mai îndepărtate. Pentru a prezice o valoare pentru orice locație nemăsurată, IDW folosește valorile măsurate care înconjoară locația de predicție. Valorile măsurate cel mai apropiate de locația de predicție au o influență mai mare asupra valorii preconizate decât cele mai depărtate. Această metodă presupune că fiecare punct măsurat are o influență locală care se diminuează odată cu distanță. Dă greutăți mai mari în punctele cele mai apropiate de locația de predicție, iar greutățile se diminuează în funcție de distanță, de unde și denumirea de distanță inversă ponderată. Greutățile alocate punctelor de date sunt ilustrate în exemplul următor:

Temă: Interpolare – Metoda ponderării inverse la distanță

Student: Belu Cosmin-Cătălin

Grupă/An: A6III

Profesor: Doru Călcai

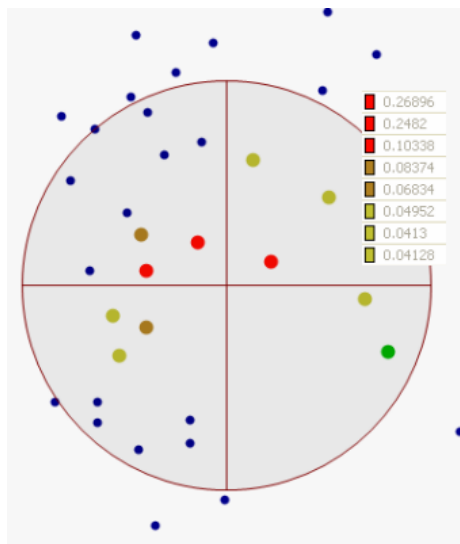


Fig.1

Funcția putere

Așa cum am menționat mai sus, greutatea este proporțională cu inversul distanței (între punctul de date și locul de predicție) ridicată la valoarea de putere p . Drept urmare, pe măsură ce distanța crește, greutatea scade rapid. Viteza la care ponderile scad este dependentă de valoarea p . Dacă $p = 0$, nu există o scădere cu distanța și, deoarece fiecare greutate este aceeași, predicția va fi media tuturor valorilor datelor din vecinătatea de căutare. Pe măsură ce p crește, ponderile punctelor îndepărtate scad rapid. Dacă valoarea p este foarte mare, numai punctele imediate din jur vor influența predicția.

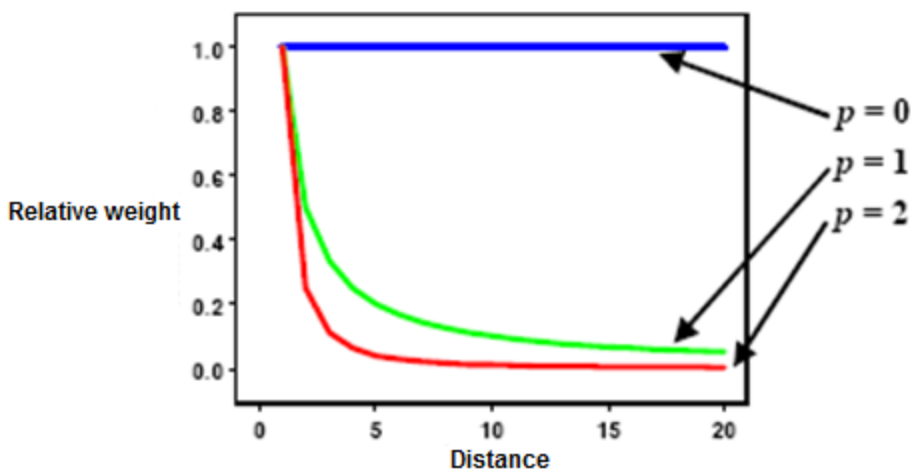


Fig.2 Scăderea greutății la distanță

Analistul geostatistic utilizează valori ale puterii mai mari sau egale cu 1. Când $p = 2$, metoda este cunoscută sub denumirea de interpolare pătrată cu distanță inversă. Valoarea implicită este $p = 2$, deși nu există nicio justificare teoretică pentru a prefera această valoare față de altele, iar efectul schimbării p ar trebui investigat prin previzualizarea ieșirii și examinarea statisticilor de validare încrucișată.

“Cartierul” de căutare

Deoarece lucrurile care se apropie unele de altele sunt mai asemănătoare cu cele care sunt mai îndepărtate, deoarece locațiile se depărtează, valorile măsurate vor avea o relație mică cu valoarea locației de predicție. Pentru a calcula viteza, puteți exclude punctele mai îndepărtate care vor avea o influență mică asupra prezicerii. Ca urmare, este o practică obișnuită să limitezi numărul de valori măsurate prin specificarea unui cartier de căutare. Forma cartierului restricționează cât de departe și unde să caute valorile măsurate pentru a fi utilizate în predicție. Alți parametri de vecinătate restricționează locațiile care vor fi utilizate în acea formă. În imaginea următoare, cinci puncte măsurate (vecini) vor fi utilizate atunci când se prezice o valoare pentru locația fără măsurare, punctul galben.

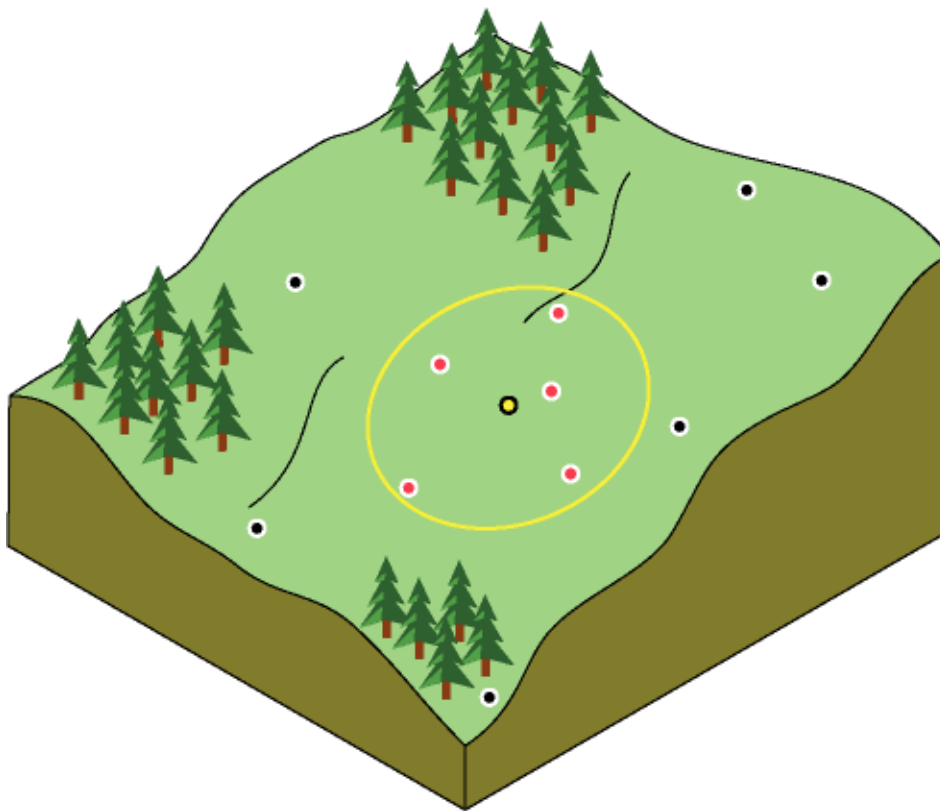


Fig.3 Exemplu

Forma cartierului este influențată de datele de intrare și de suprafața pe care încercați să le creați. Dacă nu există influențe direcționale în datele dvs., veți dori să luați în considerare punctele în mod egal în toate direcțiile. Pentru a face acest lucru, veți defini cartierul de căutare ca un cerc. Cu toate acestea, dacă există o influență direcțională în datele dvs., cum ar fi un vânt predominant, poate doriți să vă reglați schimbând forma cartierului de căutare într-o elipsă cu axa principală paralelă cu vântul. Ajustarea pentru această influență direcțională este justificată, deoarece știți că locațiile în sus dintr-o locație de predicție vor fi mai similare la distanțe îndepărtate decât locațiile care sunt perpendiculare cu vântul, dar situate mai aproape de locația de predicție.

După ce a fost specificată o formă de cartier, puteți restricționa ce locații de date din formă trebuie utilizate. Puteți defini numărul maxim și minim de locații de utilizat și puteți împărți cartierul în sectoare. Dacă împărțiți cartierul în sectoare, restricțiile maxime și minime vor fi aplicate fiecărui sector.

Punctele evidențiate în vizualizarea datelor arată locațiile și greutatea care vor fi utilizate pentru a prezice o locație în centrul elipsei (locația traversei). Cartierul de căutare este limitat la interiorul elipsei. În exemplul prezentat mai jos, celor două puncte roșii li se vor acorda greutăți mai mari de 10%. În sectorul estic, unui punct (maro) i se va acorda o pondere între 5% și 10%. Restul punctelor din cartierul de căutare vor primi greutăți mai mici.

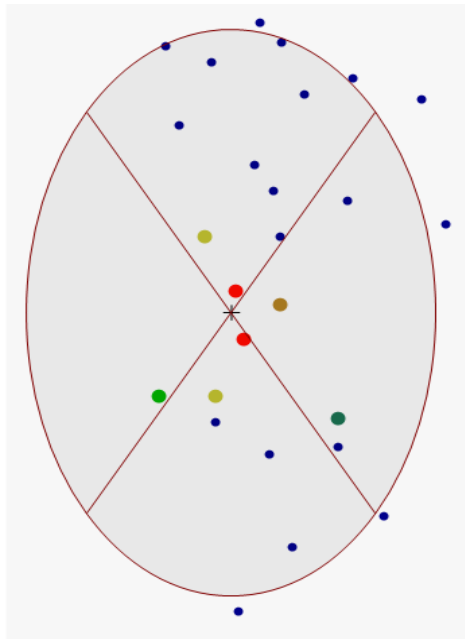


Fig.4

IV. Matematica din spatele acestei metode

Cea mai buna varianta este sa folosim un exemplu. Vom lua 3 puncte distincte astfel:

- Distanța: 350m, 750m, 850m.
- Valorile: 12, 10, 10.

Pentru o putere de 1, valoarea celulei va fi egala cu:

$$((12/350) + (10/750) + (10/850)) / ((1/350) + (1/750) + (1/850)) = 11.1$$

Pentru o putere de 2, valoarea celulei va fi egala cu:

$$((12/350^2) + (10/750^2) + (10/850^2)) / ((1/350^2) + (1/750^2) + (1/850^2)) = 11.4$$

Iar formula este următoarea:

$$z_p = \frac{\sum_{i=1}^n \left(\frac{z_i}{d_i^p} \right)}{\sum_{i=1}^n \left(\frac{1}{d_i^p} \right)}$$

Fig.5

Unde notația sigma înseamnă pur și simplu că adăugăm orice număr de puncte, ce vor fi interpolate. Se vor rezuma valorile de creștere la fiecare punct, în ceea ce privește distanță.

Un număr mai mic la numitor, înseamnă mai multă distanță și are un efect mai mic asupra valorilor interpolate. De asemenea, nu putem avea niciodată valori peste sau sub valorile maxime/minime cunoscute.

V.Concluzie

O suprafață calculată folosind IDW depinde de selectarea valorii puterii (p) și a strategiei de căutare a vecinătății. IDW este un interpolator exact, unde valorile maxime și minime pe suprafața interpolată pot apărea numai în punctele de probă.

Suprafața de ieșire este sensibilă la aglomerare și la prezența valorilor exterioare. IDW presupune că fenomenul modelat este determinat de variația locală, care poate fi capturată (modelată) prin definirea unui cartier de căutare adecvat. Deoarece IDW nu oferă erori standard de predicție, justificarea utilizării acestui model poate fi problematică.

VI.Bibliografie

<https://pro.arcgis.com/>

<https://en.wikipedia.org/>

<https://gisgeography.com/>