

Optimizare

Laborator 2 : Metoda Gradient

1 Metoda Gradient

Fie problema de optimizare neliniara fara constrangeri (UNLP):

$$x^* = \arg \min_{x \in \mathbb{R}^n} f(x) \quad (1)$$

O metoda clasica de rezolvare a acestei probleme o reprezinta **Metoda Gradient (MG)** (*engl. Gradient Descent*). Caracteristici:

- Este o metoda de ordinul 1, i.e foloseste *informatia de gradient*.
- Este un **algorithm iterativ**, i.e. gaseste o **solutie aproximativa** cu o **acuratete** prestabilita ϵ .

Pseudocodul metodei gradient:

Algoritmul MG

Date de intrare : \mathbf{x}_0 ($t = 0$) punctul initial,
pasul $\alpha_t > 0$

1. Atata timp cat $\text{criteriu}(x_t) \geq \epsilon$:

1.1 $\mathbf{x}_{t+1} = \mathbf{x}_t - \alpha_t \nabla f(\mathbf{x}_t)$

1.2 $t = t + 1$

2.Returneaza x_{t+1}

unde $\text{criteriu}(x_t)$ se numeste criteriu de oprire si poate fi:

$\|\nabla f(x_t)\|$ sau $|f(x_t) - f^*|$ sau $\|x_{t+1} - x_t\|$

Teorema (Conditii necesare de ordinul I): Fie $f \in \mathcal{C}^1$ si $x^* \in \text{dom } f$ un punct de minim local al (UNLP). Atunci avem:

$$\nabla f(x^*) = 0.$$

Drept consecinta a teoremei de sus, criteriul de oprire ideal pentru algoritmii de optimizare il reprezinta

$$\text{criteriu}(x_t) = \|\nabla f(x_t)\|.$$

Atunci cand functia obiectiv are gradient Lipschitz, pasul de descrestere constant ideal este $\alpha = \frac{1}{L}$.

2 Sisteme de recomandare

2.1 Motivatie

Consumatorii au nevoie de recomandări de încredere deoarece având la dispoziție un număr aproape nelimitat de opțiuni, aceștia doresc să își minimizeze timpul de căutare și să primească opțiuni viabile. Cu acest scop, s-au dezvoltat sisteme de recomandare.

2.2 Tipuri de sisteme de recomandare

- Bazate pe **filtru de continut** - folosește metadate pentru a determina gusturile utilizatorului. De exemplu, sistemul recomandă utilizatorului filme în funcție de preferințele lor de genuri, actori, teme etc. Un astfel de sistem se potrivește utilizatorului și articolului pe baza similitudinii.
- Bazat pe **filtru colaborativ** - aceasta strategie recomandă consumatorilor articole pe baza comportamentului lor observat.

În cele ce urmează vom formula și implementa sistemul de recomandare bazat pe filtrul colaborativ.

2.3 Formularea problemei

Fie matricea de feedback $R \in \mathbb{R}^{n \times N}$ (comportamentul observabil) unde

- n - numărul de utilizatori;
- N - numărul de articole;
- $r_{i,j}$ cu $i = 1 : n$ și $j = 1 : N$ - feedback-ul utilizatorului i pentru articolul j .

Convenție: $r_{i,j} = 0$ când utilizatorul i nu oferă feedback articolului j . Prin urmare, matricea R **este foarte rară**, dat fiind că un utilizator nu poate oferi feedback pentru toate articolele existente.

Scop: Dorim să descoperim niste caracteristici latente (i.e. neobservabile) în comportamentul utilizatorilor. Prin urmare, descompunem matricea R în produs de două matrice $U \in \mathbb{R}^{n \times k}$ și $M \in \mathbb{R}^{N \times k}$, unde

- k - numărul de caracteristici latente (i.e. neobservabile);
- $U \in \mathbb{R}^{n \times k}$ - matricea utilizatorilor
- $M \in \mathbb{R}^{N \times k}$ - matricea de filme

Stiai? Folosind o altă convenție pentru lipsa de feedback, se poate folosi DVS-ul pentru a găsi caracteristici latente folosind descompunerea lui R în $U\Sigma V^T$. U reprezintă matricea utilizatorilor și $M = \Sigma V^T$ reprezintă matricea filmelor.

Alegerea funcției obiectiv:

$$\min_{U \in \mathbb{R}^{n \times k}, M \in \mathbb{R}^{N \times k}} \frac{1}{2} \|R - UM^T\|_F^2 \Rightarrow \min_{U \in \mathbb{R}^{n \times k}, M \in \mathbb{R}^{N \times k}} f(U, M) := \frac{1}{2} \sum_{i,j \in \text{obs}} (R_{i,j} - \langle U_i, M_j \rangle)^2, \quad (2)$$

unde $\text{obs} = \{(i, j) | r_{i,j} \neq 0 \ \forall i = 1 : n, j = 1 : N\}$.



$$R_{n \times N} \approx U_{n \times k} M_{N \times k}^T$$

O data gasite U si M cu metoda gradient de exemplu, acestea se vor inmulti pentru a vedea predictia pentru articole.

Putem mai departe sa analizam similaritatea intre doi utilizatori apeland de exemplu la similaritatea cosinusoidala:

$$\cos \theta = \frac{\langle U(i, :), U(j, :) \rangle}{\|U(i, :)\| \|U(j, :)\|},$$

unde θ este unghiul dat de cei doi vectori $U(i, :)$ si $U(j, :)$. Interpretare:

- $\cos \theta \approx 1$ - utilizatorul i si j consuma acelasi continut si impartasesc aceasi parere despre continut.
- $\cos \theta \approx -1$ - utilizatorul i si j consuma acelasi continut, dar au pareri diferite despre continut.
- $\cos \theta \approx 0$ - utilizatorul i si j nu consuma acealasi continut.

3 Cerinta laborator

In cele ce urmeaza vom utiliza baza de date de mici dimensiuni de la MovieLens [2]. Aceasta contine 610 utilizatori ce au acordat ratinguri de 5 stele la 9742 de filme. Am prelucrat datele selectand doar filmele ce au primit cel putin 5 ratinguri si am selectat doar 500 de utilizatori. Astfel matricea $R_{\text{train}} \in \mathbb{R}^{500 \times 3650}$.

- (2p) Rezolvati problema (2) implementand metoda gradient pentru $\epsilon = 0.001$ si $\alpha_k = 0.003$.

Hint: Derivatele partiale pentru functia obiectiv din (2) sunt:

$$\frac{\partial f}{\partial U_{i,k}} = -(R_{i,j} - \langle U_i, M_j \rangle) M_{j,k}$$

$$\frac{\partial f}{\partial M_{j,k}} = -(R_{i,j} - \langle U_i, M_j \rangle) U_{i,k}$$

- (0.5p) Afisati intr-un grafic evolutia functiei obiectiv (folositi comanda *plot*) si intr-un alt grafic evolutia criteriului de oprire (folositi comanda *semilogy*).
- (0.5p) Calculati similaritatea cosinusoidala intre doi utilizatori la alegere. Care este interpretarea?

References

- [1] <https://developers.google.com/machine-learning/recommendation/collaborative/matrix>
- [2] <https://grouplens.org/datasets/movielens/>