Βελτιστοποίηση

Ιωάννης Γ. Τσούλος

Τμήμα Πληροφορικής και τηλεπικοινωνιών
Πανεπιστήμιο Ιωαννίνων

2022

Περίληψη

Μέθοδος διχοτόμησης

- Εστω [a,b] το πεδίο έρευνας.
- Αρχική υπόθεση: $f'(a) < 0$ και $f'(b) > 0$.
- Επιλέγεται το ενδιάμεσο σημείο:

$$c = a + \frac{b - a}{2}$$

- Αν $f'(c) > 0$ τότε γίνεται αναζήτηση στο [a,c] αλλιώς γίνεται αναζήτηση στο [c,b].
- Το αρχικό ζητούμενο είναι να βρεθεί το διάστημα [a,b].

Ο αλγόριθμος της μεθόδου διχοτόμησης

1. **Θέτουμε** k=0
2. **Αν** $f'(a) > 0$ η $f'(b) < 0$ **τερματισμός.**
3. **Επανέλαβε**
    1. **Θέτουμε** $c = a + \frac{b-a}{2}$
    2. **Αν** $f'(c) > 0$ τότε $b = c$
    3. **Αλλιώς** $a = c$
    4. k=k+1
    5. **Αν** $|b - a| \leq e$, τερματισμός όπου e μια πολύ μικρή τιμή.
4. **Τέλος - Επανάληψης.**

Παρεμβολή Hermite

- Από τα σημεία a,b διέρχεται ένα μοναδικό πολυώνυμο τρίτου βαθμού:

$$p(x) = c_1 x^3 + c_2 x^2 + c_3 x + c_4$$

- Ισχύει:
  $p(a) = f(a),\ p'(a) = f'(a),\ p(b) = f(b),\ p'(b) = f'(b)$
- Για την εύρεση των συντελεστών του πολυωνύμου απαιτείται η επίλυση του ακόλουθου γραμμικού συστήματος:

$$\begin{aligned} c_1 a^3 + c_2 a^2 + c_3 a + c_4 &= f(a) \\ c_1 b^3 + c_2 b^2 + c_3 b + c_4 &= f(b) \\ 3c_1 a^2 + 2c_2 a + c_3 &= f'(a) \\ 3c_1 b^2 + 2c_2 b + c3 &= f'(b) \end{aligned}$$

- Μπορεί να επιλυθεί με διάφορες μεθόδους όπως με Gauss.

Κυβική παρεμβολή

1. Μέθοδος Davidon.
2. Από τα σημεία $(a, f(a))$ και $(b, f(b))$ διέρχεται ένα πολυώνυμο τρίτου βαθμού p(x)
3. Θα πρέπει να ισχύει επιπλέον $p'(a) = f'(a),\ p'(b) = f'(b)$
4. Εντοπίζεται η θέση του ελαχίστου του πολυωνύμου $p(x)$
5. Το σημείο αυτό θα αντικαταστήσει είτε το a είτε το b.

Μέθοδος Davidon (αλγόριθμος)

1. **Θέσε** k=0
2. **Επανέλαβε**
    1. Υπολογισμός πολυωνύμου Hermite $H(a, b)$
    2. Εύρεση ελαχίστου

    $$x^* = \min_x H(a, b)$$

    Η εύρεση του σημείου x^* μπορεί να γίνει και με γραμμική αναζήτηση.
    3. **Αν** $f(x^*) < f(a), \ a = x^*$
    4. **Αλλιώς** $b = x^*$
    5. k=k+1
    6. **Αν** $|b - a| \leq \epsilon$, τερματισμός.
3. **Τέλος Επανάληψης**

Gradient Descent

- Στην μέθοδο αυτή γίνεται αναζήτηση του ελαχίστου κατά μήκους της $-\nabla f(x)$
- Η μέθοδος είναι γνωστή και με το όνομα Back Propagation όταν εφαρμόζεται στην ελαχιστοποίηση σφαλμάτων σε Τεχνητά νευρωνικά δίκτυα.
- Αποτελεί την βάση για πιο σύνθετες και πιο αποδοτικές τεχνικές βελτιστοποίησης.
- Βασίζεται πολύ στο βήμα h που θα επιλεγεί.

Ο αλγόριθμος Gradient Descent

1. **Θέσε** x_0
2. **Θέσε** $k = 0$
3. **Μέχρι** να ικανοποιηθεί το κριτήριο τερματισμού κάνε
    1. **Υπολόγισε** $x_{k+1} = x_k - h\nabla f\left(x_k\right)$
    2. $k = k + 1$
4. **Τέλος - Μέχρι**

Κριτήρια τερματισμού

- Μέγιστος αριθμός επαναλήψεων.
- $\|x_{k+1} - x_k\| \leq e$
- $|f(x_{k+1}) - f(x_k)| \leq e$
- $\frac{\|x_{k+1}-x_k\|}{\|x_k\|} \leq e$ (δεν δουλεύει αν $x_k \simeq 0$)

Υπολογισμός h

- Μπορεί να τεθεί σε πολύ χαμηλή τιμή (αργή σύκλιση).
- Μπορεί να ξεκινάει από υψηλές τιμές και να μειώνεται (όπως η θερμοκρασία στο Simulated Annealing).
- Μπορεί να υπολογιστεί με ελαχιστοποίηση της συνάρτησης $\phi(h) = f\left(x_k - h\nabla f\left(x_k\right)\right)$ (γραμμική αναζήτηση).
- Θέλει προσοχή ο υπολογισμός καθώς μπορεί να δημιουργηθούν και σημεία **εκτός πεδίου ορισμού** της συνάρτησης.

Χρήση Momentum στην μέθοδο Gradient Descent

- Για τον υπολογισμό του νέου σημείου λαμβάνεται υπόψιν και το ιστορικό των μεταβολών
- Εισάγεται μια ακόμα παράμετρος με το όνομα momentum (παράμετρος m)
- Η ενημέρωση γίνεται σύμφωνα με τον τύπο:

$$x_{k+1} = x_k - h\nabla f\left(x_k\right) - m\left(x_k - x_{k-1}\right)$$

- Οι μεταβολές είναι πιο ομαλές, αλλά απαιτείται η χρήση δύο παραμέτρων.

Η μέθοδος ADAM

- Χρησιμοποιείται από το 2015 ευρύτατα στα βαθιά νευρωνικά δίκτυα.
- Δεν απαιτεί την χρήση δευτέρων παραγώγων.
- Είναι χρήσιμη σε περιπτώσεις που η παράγωγος έχει θόρυβο στον υπολογισμό της ή και υπάρχουν προβλήματα αριθμητικής ακρίβειας.
- Χρησιμοποιεί διανύσματα momentum και όχι απλώς έναν αριθμό.
- Παίρνει μέρος στον υπολογισμό τόσο η παράγωγος όσο και το τετράγωνό της.

Ο αλγόριθμος της μεθόδου ADAM (παράμετροι)

- Παράμετρος α, μέγεθος βήματος, συνήθως 0.001
- Παράμετροι β_1, β_2, είναι οι παράγοντες momentum με συνήθεις τιμές 0.9 και 0.999

Ο αλγόριθμος ADAM (βήματα)

1. **Θέσε** $m_0 = 0$, πρώτο διάνυσμα momentum
2. **Θέσε** $u_0 = 0$, δεύτερο διάνυσμα momentum
3. **Θέσε** $k = 0$, αριθμός επαναλήψεων
4. **Έστω** ένα νέο δείγμα από την συνάρτηση x_0
5. **Μέχρι** τερματισμό **κάνε**
    1. $g_k = \nabla f(x_k)$
    2. $m_{k+1} = \beta_1 m_k + (1 - \beta_1) g_k$
    3. $u_{k+1} = \beta_2 u_k + (1 - \beta_2) g_k^2$
    4. $\hat{m}_{k+1} = \frac{1}{1-\beta_1^{k+1}} m_{k+1}$, παίρνει δυνάμεις για τα β_1, β_2 ώστε να μειώνεται η επίδρασή τους όσο περνάνε οι επαναλήψεις.
    5. $\hat{u}_{k+1} = \frac{1}{1-\beta_2^{k+1}} u_{k+1}$
    6. $x_{k+1} = x_k - \alpha \frac{\hat{m}_{k+1}}{\left(\sqrt{\hat{u}_{k+1}} + \epsilon\right)}$
    7. $k = k + 1$
6. **Τέλος - Μέχρι**

Παραλλαγές μεθόδου ADAM και Gradient Descent

- AdaGrad
- AdaDelta
- AdaMax
- AMSGrad
- RMSProp
- NADam

Σύνοψη

- Μέθοδοι με παραγώγους στην μία διάσταση
- Μέθοδοι με παραγώγους σε πολλές διαστάσεις

Βιβλιογραφία I

- Βελτιστοποίηση τεχνικών συστημάτων, Άγγελος Πρωτοπαπάς, 2015, Κάλλιπος, ISBN: 978-960-603-493-0
- Καθολική βελτιστοποίηση: μέθοδοι λογισμικό και εφαρμογές, Ιωάννης Γ. Τσούλος, διδακτορική διατριβή, εθνικό αρχείο διδακτορικών διατριβών.
- Αριθμητικές μέθοδοι βελτιστοποίησης, Ισαάκ Λαγαρής, ιστοσελίδα σημειώσεων διαθέσιμη από http://www.cs.uoi.gr/~lagaris/OPT_UNDER/
- Τεχνικές βελτιστοποίησης, Γεώργιος Α. Ροβιθάκης, 2007, εκδόσεις Τζιόλα, ISBN-13: 978-960-418-141-4 .

Βιβλιογραφία II

- Μαθηματική θεωρία βελτιστοποίησης , DingZhu Du, Panos M. Pardalos, Weili Wu, 2005, Εκδόσεις Νέων Τεχνολογιών, ISBN 960-8105-79-X, ISBN-13 978-960-8105-79-9
- https://arxiv.org/pdf/1412.6980.pdf Η μέθοδος ADAM