Μάθηση k-τροπικών κατανομών μέσω δοκιμών
Learning k-modal distributions via testing

Ιωσήφ Μουλίνος

Εθνικό Μετσόβιο Πολυτεχνείο

8 Ιουνίου 2017

Επιβλεπόμενη μάθηση

Μη επιβλεπόμενη μάθηση

Μάθηση κατανομής πιθανότητας $\equiv$ Εκτίμηση πυκνότητας πιθανότητας

Λαμβάνοντας δείγματα από την κατονομή p εξάγεται υποθετική κατανομή $\hat{p}$

Με υψηλή πιθανότητα $d_{TV}(p, \hat{p}) \leq \epsilon$.

$$d_{TV}(p,q) := max_{S\subseteq[n]}|p(S) - q(S)| = \frac{1}{2}\|p-q\|_1,$$

$$p(S) = \sum_{i\in S} p(i),\ [n] = 1, 2, 3, ..., n$$

Πληροφοριοθεωρητικά, για $\epsilon \geq 1/n^{\Omega(1)}$, η μάθηση μονότονης κατανομής απαιτεί $\Omega(log(n)/\epsilon^3)$ δείγματα [Birgé 1987]

Παραθέτοντας k μονότονα διαστήματα μήκους n/k, κάθε αλγόριθμος μάθησης k τροπικής κατανομής στο $[n]$, με $\epsilon \geq 1/n^{\Omega(1)}$ απαιτεί $\Omega(klog(n/k)/\epsilon^3)$

Ζητούμενο
Αλγόριθμος με πολυωνυμικό πλήθος δειγμάτων ως προς τον αριθμό των τρόπων και τον λογάριθμο *logn* και χρόνο εκτέλεσης $poly(k, log\ n, 1/\epsilon)$

Αλγόριθμος Birgé

p μη αύξουσα κατανομή στο [n].

1. Διαμέριση του [n] σε $I_1, \ldots, I_\ell$, $\ell = O(m^{1/3} \cdot (log\ n)^{2/3})$.
   Για $j \in [n]$ και $i \in I_i$, $p_f(i) = \sum_{t \in I_j} p(t)/|I_j|$
   $d_{TV}(p_f, p) = O(m^{1/3} \cdot (log\ n)^{2/3})$
2. Κατανομή p_r στο $[\ell]$ με $p_r(i) = p(I_i)$ για $i \in [\ell]$
   Με τα m δείγματα λαμβάνεται κατανομή $\hat{p}_r$ τ.ω.
   $E[d_{TV}(p_r, \hat{p}_r)] = O(\sqrt{\ell/m}) = O((log\ n/(m+1))^{1/3})$
   (ανισότητα Vapnik-Chervonenkis)
3. Εξάγεται η κατανομή $(\hat{p}_r)_f$ με στήριγμα το [n]
   Από βήματα 1 και 2:
   $E[d_{TV}((\hat{p}_r)_f, p_f)] = O(\sqrt{\ell/m}) = O((log\ n/(m+1))^{1/3})$
4. Από βήματα 1 και 3:
   $E[d_{TV}((\hat{p}_r)_f, p)] = O(\sqrt{\ell/m}) = O((log\ n/(m+1))^{1/3})$

Μάθηση σχεδόν μονότονων κατανομών

$\delta := O((log\ n/(m+1))^{1/3})$
Στα βήματα 2 και 3 μαθαίνεται η κατανομή p_f σε απόσταση διακύμανσης δ
Ισχυρισμός: Αν p είναι τ-κοντά σε μη αύξουσα κατανομή $p^{\downarrow}$ τότε η p_f ικανοποιεί $d_{TV}(p_f, p) \leq (2\tau + \delta)$

Απόδειξη: $\tau_j = \|p - p^{\downarrow}\|_1$ στο διάστημα I_j, τότε $\sum_{i=1}^{\ell} \tau_j \leq \tau$
Birgé $\Rightarrow d_{TV}((p^{\downarrow})_f, p^{\downarrow}) = \delta$ άρα $d_{TV}((p^{\downarrow})_f, p) = \delta + \tau$ (1)

Σε κάθε διάστημα I_j, $\|p(I_j) - p^{\downarrow}(I_j)\|_1 \leq \tau_j$ και οι $p_f, (p^{\downarrow})_f$ ομοιόμορφες.
$\Rightarrow d_{TV}(p_f, (p^{\downarrow})_f) \leq \tau$ (2)

(1),(2), τριγωνική ανισότητα $\Rightarrow d_{TV}(p_f, p) \leq 2\tau + \delta$
Ο αλγόριθμος Birgé μαθαίνει σχεδόν μονότονη κατανομή με απόσταση διακύμανσης $2\tau + 2\delta$

Μάθηση k-τροπικών κατανομών

Αλγόριθμος Birgé σε $\binom{n}{k}$ των πιθανών ακρότατων σημείων
Καλή δειγματική πολυπλοκότητα, μη αποδοτική χρονική $\Omega(n^k)$

Διαμέριση του $[n]$ σε k/ϵ διαστήματα με χονδρικά ϵ/k πιθανότητα το καθένα

- k μη μονότονα διαστήματα
- συνολικό σφάλμα $O(\epsilon)$
- καλή χρονική πολυπλοκότητα $poly(k, log\ n, 1/\epsilon)$
- μεγαλύτερη της επιθυμιτής δειγματική,
- $\Omega(k/\epsilon)$ φορές Birgé με $\Omega(log(n)/\epsilon^3) \Rightarrow \Omega(k(log(n)/\epsilon^4)$

Αλγόριθμος ελέγχου μονοτονίας k-τροπικής κατανομής

Ο αλγόριθμος $T^{\uparrow}(\epsilon, \delta)$ ($T^{\downarrow}(\epsilon, \delta)$) χρησιμοποιεί $O(log(1/\delta)) \cdot (k/\epsilon^2)$ δείγματα από k-τροπική κατανομή.

Απαιτήσεις

(Πληρότητα) Εάν η p είναι μη φθίνουσα (μη αύξουσα), τότε ο $T^{\uparrow}$ ($T^{\downarrow}$) δίνει 'ναι' με πιθανότητα τουλάχιστον $1 - \delta$.

(Συνέπεια) Εάν η p είναι ϵ-μακρυά από το να είναι μη αύξουσα (μη φθίνουσα), τότε ο $T^{\uparrow}$ ($T^{\downarrow}$) δίνει 'ναι' με πιθανότητα τω πολύ δ.

Ο αλγόριθμος $T^{\uparrow}(\epsilon, \delta)$ υλοποιείται επαναλαμβάνοντας τον αλγόριθμο $T^{\uparrow}$, $O(k/\tau^2)$ φορές.

Δοκιμαστής $T^{\uparrow}(\tau)$

Είσοδος: $\tau > 0$· δείγματα της k-τροπικής κατανομής q στο $[n]$

1. Λήψη $\Theta(k/\tau^2)$ δειγμάτων **s** από την q στο $[n]$
2. Εάν υπάρχει $\ell \in [k]$ και $\{\alpha_i, b_i, c_i\}_{i=1}^{\ell} \in s \cup n$, με $\alpha_i \leq b_i < c_i < \alpha_{i+1}$, $i \in [\ell - 1]$, τ.ω. $\sum_{i=1}^{\ell} T(\hat{q}, \alpha_i, b_i, c_i - 1) \geq \tau/4$ τότε απάντησε ¨όχι¨, άλλως απάντησε ¨ναι¨.

Θεώρημα

Ο αλγόριθμος $T^{\uparrow}$ χρησιμοποιεί $O(k/\tau^2)$ δείγματα της q, εκτελεί $poly(k/\tau) \cdot logn$ διφυοπράξεις και ικανοποιεί τις απαιτήσεις πληρότητας και συνέπειας.

Λήμμα

Έστω q k-τροπική κατανομή στο $[n]$ που είναι τ μακρυά από το να είναι μη φθίνουσα. Υπάρχει $\ell \in [k]$ και $\{\alpha_i, b_i, c_i\}_{i=1}^{\ell} \subseteq [n]^{3\ell}$ με $\alpha_i \leq b_i < c_i < \alpha_{i+1}$, $i \in [\ell - 1]$, τ.ω.
$\sum_{i=1}^{\ell} T(q, \alpha_i, b_i, c_i) \geq \tau/2$

Απόδειξη λήμματος με αντιθετοαντιστροφή, μειώνοντας βαθμηδόν κατά 1 την τροπικότητα της κατανομής, κρατώντας το σφάλμα ελεγχόμενο (θυμίζει Myerson ironing).

Έλεγχος ύπαρξης τριάδων $\{\alpha_i, b_i, c_i\}_{i=1}^{\ell} \in s \cup n$, με $\alpha_i \leq b_i < c_i < \alpha_{i+1}$, $i \in [\ell - 1]$, τ.ω. $\sum_{i=1}^{\ell} T(\hat{q}, \alpha_i, b_i, c_i - 1) \geq \tau/4$.
Με ωμή δύναμη $\Omega(r^k) \cdot logn$.

Θεωρούμε $T(\ell) = max\{\sum_{i=1}^{\ell} T(\hat{q}, \alpha_i, b_i, c_i - 1) | \{\alpha_i, b_i, c_i\}_{i=1}^{\ell} \in s \cup n \textit{ with } \alpha_i \leq b_i < c_i < \alpha_{i+1}, i \in [\ell - 1]\}$
Με δυναμικό προγραμματισμό

1. Δειγματοληψία της p σε $r = \Theta(1/\tau^2)$ θέσεις του στηρίγματος, $\tau := \epsilon/(100k)$.
2. Διαμέριση του $[n]$ άπληστα σε ℓ διαστήματα με $\mathcal{I} = \{I_i\}_{i=1}^{\ell}$
   1. $I := [1, j_1]$ με $j_1 := min\{j \in [n] \mid \hat{p}([1, j]) \geq \frac{\epsilon}{10k}\}$
   2. Για $i \geq 1$ εάν $\cup_{j=1}^{i} I_j = [1, j_i]$ τότε $I_{i+1} := [j_i + 1, j_{i+1}]$, όπου
      εάν $\hat{p}([j_i + 1, n]) \geq \epsilon/(10k)$,
      $j_{i+1} := min\{j \in [n] \mid \hat{p}([j_i + 1, j]) \geq \epsilon/(10k)\}$
      άλλως, $j_{i+1} := n$
3. $\tau' := \epsilon/(2000k)$. Λήψη
   $r' = \Theta((k^2/\epsilon^3) \cdot log(1/\tau')loglog(1/\tau'))$ δειγμάτων s της p
4. Ταυτόχρονη εκτέλεση των αλγορίθμων $T^{\uparrow}(\epsilon, \tau')$ και $T^{\downarrow}(\epsilon, \tau')$ στις μερικές κατανομές $p_{\cup_{i=1}^{j} I_i}$ για $j = 1, 2, ...,$ έως εύρεση I_{j_1} για το οποίο οι $T^{\uparrow}(\epsilon, \tau')$ και $T^{\downarrow}(\epsilon, \tau')$ επιστρέψουν "όχι" στο $p_{\cup_{i=1}^{j_1}} I_i$.
   $I_{j_1} = [\alpha_{j_1}, b_{j_1}]$ διακρίνονται 2 περιπτώσεις

Εάν $\hat{p}[\alpha_{j_1}, b_{j_1}] \geq 2\epsilon/(10k)$, όρισε $I'_{j_1} := [\alpha_{j_1}, b_{j_1} - 1]$ και το b_{j_1} είναι βαρύ σημείο.
Εάν $\hat{p}[\alpha_{j_1}, b_{j_1}] < 2\epsilon/(10k)$, όρισε $I'_{j_1} := I_{j_1}$.

Ονόμασε το I'_{j_1} αμελητέο διάστημα. Εάν $j_1 > 1$, όρισε το πρώτο υπερδιάστημα να είναι $\cup_{i=1}^{j_1-1} I_i$ και θέσε το $\alpha_1 \in \{\uparrow, \downarrow\}$ σε $\uparrow$ εάν $T\uparrow$ επέστρεψε "ναι" στο $p_{\cup_{i=1}^{j_1-1} I_i}$ και θέσε $\alpha_1 =\downarrow$ εάν $T\downarrow$ επέστρεψε "ναι" στο $p_{\cup_{i=1}^{j_1-1} I_i}$.

5. Επανάλαβε το βήμα 4 ξεκινώντας από το επόμενο διάστημα I_{j_1+1} βρες το αριστερότερο διάστημα I_{j_2} για το οποίο αμφότερα τα $T^{\uparrow}$ και $T^{\downarrow}$ επιστρέφουν "όχι" στο $p_{\cup_{i=1}^{j_1} I_i}$. Συνέχισε μέχρί όλα τα διαστήματα ως το I_ℓ να έχουν χρησιμοποιηθεί. Έστω $S_1, ..., S_t$ τα υπερδιαστήματα που λαμβάνονται μέσω της παραπάνω διαδικασίας και $\alpha_1, ..., \alpha_t \in \{\uparrow, \downarrow\}^t$ η ακολουθία μονοτονίας τους.

6. Λάβε $m = \Theta(k \cdot log(n/k)/\epsilon^3)$ δείγματα s' από την p. Για κάθε υπερδιάστημα $S_i, i \in [t]$, τρέξε τον αλγόριθμο A^{α_i} στην υπό συνθήκη κατανομή p_{S_i} της p, χρησιμοποιώντας δείγματα από το $s' \cup S_i$. Έστω $\tilde{p}_{S_i}$ η υπόθεση που λαμβάνεται.
7. Δώσε την υπόθεση $h = \sum_{i=1}^{t} \hat{p}(S_i) \cdot \tilde{p}_{S_i} + \sum_j \hat{p}(\{b_j\}) \cdot 1_{b_j})$.

Έλεγχος υπόθεσης

Θεώρημα

Υπάρχει αλγόριθμος Έλεγχος-Υπόθεσης$^{p}(h_1, h_2, \epsilon', \delta')$ που δεδομένων δύο υποθέσεων h_1, h_2 για την p, παράμετρο ακρίβειας ϵ' και παράμετρο εμπιστοσύνης δ', λαμβάνει $m = O(log(1/\delta')/\epsilon'^2)$ δείγματα από την p και επιστρέφει υπόθεση h $\in \{h_1, h_2\}$. Εάν μία εκ των h_1, h_2 έχει $d_{TV}(h_i, p) \leq \epsilon'$, τότε με πιθανότητα $1 - \delta'$ η υπόθεση h που επιστρέφει ο αλγόριθμος Έλεγχος-Υπόθεσης έχει $d_{TV}(h, p) \leq 6\epsilon'$.

Απόδειξη

$\mathcal{W}$ στήριγμα του p,
$\mathcal{W}_1 = \mathcal{W}_1(h_1, h_2) := \{w \in \mathcal{W} \mid h_1(w) > h_2(w)\}$
$p_1 = h_1(\mathcal{W}_1), q_1 = h_2(\mathcal{W}_1)$
$p_1 > q_1$ και $d_{TV}(h_1, h_2) = p_1 - q_1$

1. αν $p_1 - q_1 \leq 5\epsilon'$ ισοπαλία, δώσε κάποιο h_i, άλλως
2. $m = O(\frac{log(1/\delta')}{\epsilon'^2})$ δείγματα $s_1, ...s_m$ από την p,
   $\tau = \frac{1}{m}|\{i|s_i \in \mathcal{W}_1\}$
3. αν $\tau > p_1 - \frac{3}{2}\epsilon'$ επίστρεψε h_1 νικητή, άλλως
4. αν $\tau < q_1 + \frac{3}{2}\epsilon'$ επίστρεψε h_2 νικητή, άλλως
5. ισοπαλία, δώσε κάποιο h_i

Εάν $d_{TV}(p, h_1) \leq \epsilon'$

αν $d_{TV}(p, h_2) > 6\epsilon' \Rightarrow$ πιθανότητα ο h_1 να μην είναι νικητής $\leq e^{-m\epsilon'^2/2}$

αν $d_{TV}(p, h_2) > 4\epsilon' \Rightarrow$ πιθανότητα ο h_2 νικητής $\leq e^{-m\epsilon'^2/2}$

απόδειξη

$r = p(W_1)$, τότε $|r - p_1| \leq \epsilon'$

Ορίζουμε $\{Z_j\}_{j=1}^{m}$ δείκτριες τυχαίες συναρτήσεις με $Z_j = 1$ ανν $s_j \in \mathcal{W}_1$, τότε $\tau = \frac{1}{m}\sum_{j=1}^{m} Z_j$

$E[\tau] = E[Z_j] = r$ Z_j ανεξάρτητες άρα από φράγμα Chernoff

$Pr[\tau \leq r - \epsilon'/2] \leq e^{-m\epsilon'^2/2} \Rightarrow Pr[\tau \leq p_1 - 3\epsilon'/2] \leq e^{-m\epsilon'^2/2}$

αν $d_{TV}(p, h_2) > 6\epsilon'$, απο τριγ. ανισότητα $p_1 - q_1 = d_{TV}(h_1, h_2) > 5\epsilon'$ και θα σταματήσουμε στο βήμα 3 με πιθανότητα $> 1 - e^{-m\epsilon'^2/2}$

αν $p_1 - q_1 \leq 5\epsilon'$ ισοπαλία, άλλως $p_1 - q_1 > 5\epsilon'$ και ο h_2 ανακηρύσσεται νικητής με πιθανότητα τω πολύ $e^{-m\epsilon'^2/2}$

Έστω ότι έχουμε αλγόριθμο που με εμπιστοσύνη 9/10 μαθαίνει την k-τροπική κατανομή με ακρίβεια ϵ. Για να ωθήσουμε την πιθανότητα σε $1-\delta$, τρέχουμε τον αλγόριθμο $O(log(1/\delta))$ φορές και χρησιμοποιείται ο Έλεγχος-Υπόθεσης σε κατάλληλο διαγωνισμό για επιλογή της κατάλληλης υπόθεσης.

Λήμμα

Έστω p τυχαία κατανομή επί του $\mathcal{W}$. Έστω D_ϵ συλλογή N κατανομών στο $\mathcal{W}$ τ.ω. υπάρχει $q \in D_\epsilon$ με $d_{TV}(p,q) \leq \epsilon$. Τότε υπάρχει αλγόριθμος που χρησιμοποιεί $O(\epsilon^{-2} logN \cdot log(1/\delta))$ δείγματα από την p και με πιθανότητα $1-\delta$ εξάγει κατανομή $p' \in D_\epsilon$ που ικανοποιεί την $d_{TV}(p,p') \leq 6\epsilon'$.

Εκτελούνται $O(log(1/\delta))$ επαναλήψεις αλγορίθμου μάθησης που παράγει ϵ-ακριβή υπόθεση με πιθανότητα $\geq 9/10$.
Τότε με πιθανότητα αποτυχίας τω πολύ $\delta/2$, τουλάχιστον μία από τις υποθέσεις που παράγονται είναι ϵ κοντά στην πραγματική κατανομή που ικανοποιεί τις προϋποθέσεις του λήμματος.

Με $O(\epsilon^{-2} logN \cdot log(1/\delta))$ δείγματα μαθαίνουμε με ακρίβεια 6ϵ και εμπιστοσύνη $1 - \delta/2$.

Συνολική δειγματική πολυπλοκότητα:
$O(log(1/\delta))$ x πολυπλ. αλγ. μάθησης $+ \ O(\epsilon^{-2} logN \cdot log(1/\delta))$.

Κεντρικό θεώρημα

Έστω p οποιαδήποτε άγνωστη κατανομή k τρόπων στο [n].

Υπάρχει αλγόριθμος που χρησιμοποιεί

$\left(\frac{k \cdot log(n/k)}{\epsilon^3} + \frac{k^2}{\epsilon^3} log(\frac{k}{\epsilon}) \cdot loglog\frac{k}{\epsilon}\right) \cdot \tilde{O}(log\frac{1}{\delta})$ δείγματα της p, εκτελεί

$poly(k, log\ n, 1/\epsilon, log(1/\delta))$ διφυοπράξεις και με πιθανότητα

$1-\delta$ δίνει υπόθεση $\hat{p}$ στο $[n]$ τ.ω. η απόσταση ολικής

διακύμανσης μεταξύ των p και $\hat{p}$ είναι τω πολύ ϵ.

($\tilde{O}$: πολυλογαριθμικός όρος. Π.χ. ως $\tilde{O}(alogb)$ συμβολίζεται

όρος της μορφής $O((alogb) \cdot (log(alogb))^c)$)