

1 Εκτίμηση Κατάταξης

Ας υποθέσουμε ότι έχουμε μια αταξινόμητη λίστα διακριτών αριθμών $x_1, x_2, \dots, x_n$. Θέλουμε να επιλέξουμε τυχαία $m < n$ στοιχεία από αυτήν τη λίστα (με αντικατάσταση) και με βάση αυτό το δείγμα θέλουμε να επιστρέψουμε κάποιο στοιχείο x έτσι ώστε το $\text{rank}(x)$ είναι περίπου ίσο με k για ένα δεδομένο k . Με τον όρο rank εννοούμε την κατάταξη του x στην αρχική λίστα. Το μεγαλύτερο από τα $x_1, x_2, \dots, x_n$ έχει rank 1, η δεύτερη μεγαλύτερη έχει rank 2 κ.λ.π. και προσπαθούμε να βρούμε κάτι με $\text{rank}(x) \approx k$. Περιγράψτε μια απλή στρατηγική για την επιλογή ενός τέτοιου x .

Πόσο μεγάλο πρέπει να θέσουμε το m (με συμβολισμό big-O) έτσι ώστε, με πιθανότητα $(1 - \delta)$, η στρατηγική σας να επιστρέφει ένα x με $(1 - \varepsilon)k \leq \text{rank}(x) \leq (1 + \varepsilon)k$;

2 Καταμέτρηση Μέσω Ερωτημάτων

Ο φίλος σας έχει στο μυαλό του ένα σύνολο αριθμών $S \subseteq \{1, \dots, n\}$. Εάν τον ρωτήσετε για ένα σύνολο $Q \subseteq \{1, \dots, n\}$ θα απαντήσει αν η τομή $S \cap Q$ είναι άδεια ή όχι. Ο στόχος σας είναι να υπολογίσετε (περίπου) πόσα στοιχεία περιέχονται στο σύνολο S .

- (α) Σχεδιάστε μια στρατηγική που να διακρίνει εάν το S περιέχει $\leq k$ ή $\geq (1 + \varepsilon)k$ για οποιοδήποτε $k \in \{1, \dots, n\}$. Δείξτε ότι $O(\log(1/\delta)/\varepsilon^2)$ ερωτήματα επαρκούν για να διακριθούν οι 2 περιπτώσεις με πιθανότητα $1 - \delta$.

Συμβουλή: Επιλέξτε ένα τυχαίο σύνολο Q που περιέχει κάθε στοιχείο $i \in \{1, \dots, n\}$ με πιθανότητα $1/k$.

- (β) Δείξτε ότι αυτό συνεπάγεται έναν αποδοτικό εκτιμητή $\hat{s}$ για το μέγεθος $|S|$, που με πιθανότητα τουλάχιστον $2/3$ ικανοποιεί $\hat{s} \leq |S| \leq (1 + \varepsilon)\hat{s}$ χρησιμοποιώντας την προαναφερθείσα στρατηγική. Ο αριθμός των ερωτημάτων θα πρέπει να είναι το πολύ λογαριθμικός στο n .

- (γ) Υποθέτοντας ότι γνωρίζετε το μέγεθος $|S|$, σχεδιάστε ένα αποδοτικό σχήμα που δειγματοληπτεί ένα ομοιόμορφα τυχαίο στοιχείο από το S ρωτώντας τον φίλο σας πολλές φορές με διαφορετικά σύνολα Q . Ο αναμενόμενος αριθμός ερωτημάτων θα πρέπει να είναι το πολύ λογαριθμικός στο n .

Συμβουλή: Μπορείτε να βρείτε ένα τυχαίο σύνολο Q που περιέχει ακριβώς ένα στοιχείο από το S ; Μπορείτε να σκεφτείτε μια αποδοτική μέθοδο για να ελέγξετε εάν το Q ικανοποιεί αυτήν την ιδιότητα, και εάν την ικανοποιεί, να επιστρέψετε το μοναδικό στοιχείο του $S \cap Q$;

3 Απόσταση Kolmogorov

Η απόσταση Kolmogorov μεταξύ των συναρτήσεων μάζας πιθανότητας $p, q : [n] \rightarrow [0, 1]$ ορίζεται ως $d_K(p, q) \triangleq \max_{i \in [n]} |p([1, i]) - q([1, i])|$, όπου $p([1, u]) \triangleq \sum_{k=1}^u p(k)$.

- (α) Δώστε έναν αλγόριθμο για να μάθετε μια αυθαίρετη κατανομή στο $[n]$ σε Kolmogorov απόσταση ε με πιθανότητα $1 - \delta$ χρησιμοποιώντας $O(\log(1/\varepsilon\delta)/\varepsilon^2)$ δείγματα. Πως εξηγείτε το γεγονός ότι ο αλγόριθμός σας έχει δειγματική πολυπλοκότητα ανεξάρτητη από το n ;

Extra Credit: Μπορείτε να αφαιρέσετε τον όρο $\log(1/\varepsilon)$ για να πάρετε την βέλτιστη πολυπλοκότητα;

- (β) Δείξτε ότι οποιοσδήποτε αλγόριθμος εκμάθησης για το παραπάνω πρόβλημα απαιτεί $\Omega(\log(1/\delta)/\varepsilon^2)$ δείγματα.
- (γ) Έστω $p : [n] \rightarrow [0, 1]$ μια συνάρτηση μάζας πιθανότητας που γνωρίζουμε ότι είναι μονότονη (μη αύξουσα) στο διατεταγμένο πεδίο της, δηλ. $p_{i+1} \leq p_i$ για όλα τα $i \in [n - 1]$. Σχεδιάστε έναν αλγόριθμο που διακρίνει με πιθανότητα τουλάχιστον $1 - \delta$ τις περιπτώσεις που η p είναι ομοιόμορφη, δηλ $p = U_n$, και $d_{TV}(p, U_n) \geq \varepsilon$. Δείξτε ότι ο αλγόριθμός σας είναι βέλτιστος ως προς τη δειγματική του πολυπλοκότητα.

4 Εκμάθηση Ορθογώνιων

Η κλάση H των ορθογώνιων διαχωριστών σε d διαστάσεις περιέχει συναρτήσεις της μορφής:

$$h_{a,b}(x) = \begin{cases} +1 & \text{Αν για κάθε } i \in [d], \text{ ισχύει } x_i \in [a_i, b_i] \\ -1 & \text{διαφορετικά} \end{cases}$$

για παραμέτρους $a_1, \dots, a_d, b_1, \dots, b_d \in \mathbb{R}$. Δηλαδή, οι συναρτήσεις παίρνουν την τιμή $+1$ αν το σημείο είναι μέσα στο ορθογώνιο και -1 αλλιώς.

- (α) Να δείξετε ότι στη realizable περίπτωση, ο αλγόριθμος A που για ένα σύνολο δειγμάτων βρίσκει το μικρότερο ορθογώνιο που περιέχει τα θετικά σημεία είναι ERM.

- (β) Να δείξετε ότι αν ο A είναι (ε, δ) -realizable PAC learner με $m_H^{\text{PAC}} = O\left(\frac{d}{\varepsilon} \log(d/\delta)\right)$.

Συμβουλή: Δείξτε ότι ο αλγόριθμος A επιστρέφει ένα ορθογώνιο με παραμέτρους $\hat{a}, \hat{b}$ που είναι υποσύνολο του πραγματικού ορθογώνιου h_{a^*, b^*} . Στη συνέχεια δείξτε ότι για κάθε $i \in [d]$, $\Pr[x_i \in (a_i^*, \hat{a}_i)] < \varepsilon/(2d)$ και $\Pr[x_i \in (\hat{b}_i, b_i^*)] < \varepsilon/(2d)$ με μεγάλη πιθανότητα.

- (γ) Υπολογίστε το VC dimension της κλάσης H .

- (δ) Ποια είναι τα bounds της δειγματικής πολυπλοκότητας για realizable και agnostic PAC learning που συνεπάγονται για τον αλγόριθμο A ;

5 Εκμάθηση από 2 πηγές δειγμάτων

Εξετάστε μια παραλλαγή του μοντέλου PAC στην οποία υπάρχουν δύο πηγές δειγμάτων: η μια δίνει θετικά παραδείγματα και η άλλη αρνητικά, και τα δύο σύμφωνα με μια υποκείμενη κατανομή D στο X .

Πιο συγκεκριμένα, για μια συνάρτηση $f : X \rightarrow \{-1, +1\}$, έστω D_+ η κατανομή πάνω στο $X_+ = \{x \in X : f(x) = +1\}$ με μάζα πιθανότητας $D_+(A) = D(A)/D(X_+)$, για κάθε $A \subseteq X_+$. Ομοίως, D_- είναι η κατανομή στο $X_- = \{x \in X : f(x) = -1\}$. Ο ορισμός του PAC learnability στο μοντέλο δύο πηγών είναι ο ίδιος με τον standard ορισμό του PAC learnability εκτός από το ότι εδώ ο εκπαιδευόμενος έχει πρόσβαση σε $m_+(\varepsilon, \delta)$ i.i.d. παραδείγματα από τα D_+ και $m_-(\varepsilon, \delta)$ i.i.d. παραδείγματα από το D_- . Ο στόχος είναι να προσδιοριστεί μια h τ.ω. με πιθανότητα τουλάχιστον $1 - \delta$, $\Pr_{x \sim D_+}[h(x) \neq +1] \leq \varepsilon$ και $\Pr_{x \sim D_-}[h(x) \neq -1] \leq \varepsilon$.

- (α) Δείξτε ότι εάν μια κλάση υποθέσεων H είναι PAC learnable στο τυπικό μοντέλο μιας πηγής παραδειγμάτων, τότε είναι PAC learnable και στο μοντέλο δύο πηγών.
- (β) Έστω h_+ και h_- οι συναρτήσεις που είναι πάντα $+1$ και πάντα -1 αντίστοιχα. Υποθέτοντας ότι $h_+, h_- \in H$, δείξτε ότι εάν η κλάση H είναι PAC learnable στο μοντέλο δύο πηγών παραδειγμάτων, τότε είναι PAC learnable και στο τυπικό μοντέλο μιας πηγής.

6 SGD για Truncated Statistics

Έστω μια μονοδιάστατη κανονική κατανομή $N(\mu^*, 1)$ με άγνωστη μέση τιμή μ^* και διακύμανση 1. Παρατηρούμε δείγματα $x \sim N(\mu^*, 1)$ όπου μας δίνονται μόνο εφόσον $x \geq 0$. Έστω $N_{\geq 0}(\mu^*, 1)$ η προκύπτουσα κατανομή. Αν είναι γνωστό ότι $|\mu^*| \leq B$, δείξτε ότι αν κάνουμε SGD στην συνάρτηση $f(\mu) = \mathbf{E}_{x \sim N_{\geq 0}(\mu^*, 1)} \left[-\ln \frac{e^{-\frac{1}{2}(x-\mu)^2}}{\int_0^\infty e^{-\frac{1}{2}(z-\mu)^2} dz} \right]$ με projection στο σύνολο $[-B, B]$ μπορούμε να εκτιμήσουμε το μ^* σε απόσταση ε . Πόσα δείγματα χρειαζόμαστε;

7 Κάτω φράγμα για τον Perceptron

Δεδομένου ενός αριθμού m , βρείτε μια ακολουθία από labeled σημεία $(x_1, y_1), \dots, (x_m, y_m) \in \mathbb{R}^3 \times \{-1, +1\}$ για την οποία ο αλγόριθμος Perceptron να κάνει λάθος πρόβλεψη για κάθε σημείο ενώ το άνω φράγμα για τα λάθη που δείξαμε στην τάξη να είναι m .

Υπόδειξη: Ορίστε κάθε x_i να είναι της μορφής (a, b, y_i) , όπου $a^2 + b^2 = m^2 - 1$. Έστω w^* το διάνυσμα $(0, 0, 1)$. Τώρα, εξετάστε την απόδειξη του άνω φράγματος του Perceptron, βρείτε πού χρησιμοποιήθηκαν ανισότητες ($\leq$) αντί για ισότητες ($=$), και βρείτε σενάρια όπου η ανισότητα ισχύει πραγματικά με ισότητα.