Khớp tuyến tính từng đoạn lồi

Tóm tắt Chúng ta xét bài toán khớp hàm tuyến tính từng đoạn lồi với một số dạng cụ thể của các dữ liệu nhiều chiều cho trước. Ngoại trừ một số trường hợp đặc biệt, bài toán này khó có thể giải chính xác, vì vậy chúng ta tập trung vào các phương pháp tìm tòi tìm khớp tối ưu cục bộ. Phương pháp mà chúng ta mô tả biến đổi trên giải thuật K-mean đối với việc gom cụm dữ liệu, có thể áp dụng tốt trong thực tế, ít nhất là đối với các dữ liệu có thể khớp tốt bằng hàm lồi. Chúng ta tập trung vào dạng hàm đơn giản nhất, cực đại của một số hàm affine xác định, và sau đó chỉ ra phương pháp này có thể mở rộng cho dạng tổng quát hơn như thế nào.

1.Bài toán khớp tuyến tính từng đoạn lồi

Chúng ta xét bài toán khớp một số dữ liệu cho trước

…………………………………………

Với hàm tuyến tính từng đoạn lồi $f: \boldsymbol{R}^n \to \boldsymbol{R}$ từ tập hợp $\mathcal{F}$ các hàm được chọn. Theo tiêu chuẩn khớp bình phương tối thiểu, chúng ta thu được bài toán

Cực tiểu hóa ……………………..

Thõa mãn……………….

với biến f. Chúng ta gọi $(\boldsymbol{J}(f)/m)^{1/2}$ là khớp căn quân phương (RMS) của hàm f của dữ liệu. Bài toán khớp tuyến tính từng đoạn lồi (1) là tìm hàm f, từ tập hợp $\mathcal{F}$ cho trước của các hàm tuyến tính từng đoạn lồi, sao cho khớp RMS của các dữ liệu cho trước tốt nhất (nhỏ nhất).

Chúng ta sẽ xét trường hợp n (số chiều của dữ liệu) tương đối nhỏ, chẳng hạn không lớn hơn 5 hoặc tương tự, trong khi đó m (số điểm dữ liệu) có thể tương đối lớn, chẳng hạn 10^4 hoặc hơn. Tuy nhiên, phương pháp mà chúng ta mô tả cũng có thể áp dụng cho bất kì giá trị nào của n và m.

Một số trường hợp đặc biệt của bài toán khớp tuyến tính từng đoạn lồi (1) có thể được giải chính xác. Khi $\mathcal{F}$ bao gồm các hàm affine, tức là, f có dạng $f(x) = a^T x + b$, bài toán (1) rút về bài toán bình phương tối thiểu tuyến tính bình thường trong các tham số hàm $a\epsilon\boldsymbol{R}^n$ và $b\epsilon\boldsymbol{R}$ và đã được giải rồi. Xét một ví dụ quan trọng, trong đó $\mathcal{F}$ bao gồm tất cả các hàm tuyến tính từng đoạn từ $\boldsymbol{R}^n$ vào $\boldsymbol{R}$, và không có ràng buộc khác về dạng của f. Đây là bài toán khớp tuyến tính từng đoạn lồi không tham số. Do đó, bài toán (1) có thể được giải chính xác bằng quy hoạch bậc hai (QP); xem (Boyd và Vandenberghe 2004, Phần.6.5.5). Tuy nhiên, phương pháp không tham số này tiềm ẩn hai nhược điểm trong thực tế. Thứ nhất, QP phải được giải rất lớn (chứa hơn mn biến), chỉ giới hạn ở giá trị m vừa phải (chẳng hạn một nghìn). Nhược điểm tiềm ẩn thứ hai là khớp hàm tuyến tính từng đoạn thu được có thể rất phức tạp, với nhiều số hạng (lên đến m).

Tất nhiên, không phải tất cả các dữ liệu đều được khớp tốt (tức là, với khớp RMS nhỏ) với hàm tuyến tính từng đoạn lồi. Ví dụ, nếu dữ liệu là các mẫu từ hàm có độ cong âm (lõm), thì không thể có hàm lồi nào có thể khớp tốt nó. Hơn nữa, khớp tốt nhất (không hoàn hảo) sẽ thu được với một hàm affine. Chúng ta cũng có thể có trường hợp ngược lại: có thể là các dữ liệu có thể được khớp hoàn hảo bởi một hàm affine, tức là chúng ta có thể có $J = 0$. Trong trường hợp này chúng ta nói rằng dữ liệu được nội suy bởi hàm tuyến tính từng đoạn lồi f.

1.1 Hàm affine cực đại

Trong bài báo này chúng ta xét bài toán khớp tham số, trong đó các hàm được chọn được tham số hóa bởi một vector các hệ số có chiều xác định $\alpha\epsilon\boldsymbol{R}^p$, ở đây p là số các tham số cần thiết để mô tả các hàm được chọn. Một dạng rất đơn giản là $\mathcal{F}_{ma}^k$, tập hợp các hàm trên $\boldsymbol{R}^n$ có dạng

……………………………………………………

Tức là, cực đại của k hàm affine. Chúng ta gọi hàm dạng này là "affine cực đại" với k số hạng. Tập hợp $\mathcal{F}_{ma}^k$ được tham số hóa bởi vector hệ số

………………………………………………..

Thực sự, bất kì một hàm tuyến tính từng đoạn lồi nào trên $\boldsymbol{R}^n$ cũng có thể được biểu diễn như hàm affine cực đại, đối với k nào đó, vì vậy dạng này là phổ biến. Tuy nhiên, ở đây chúng ta quan tâm đến trường hợp số số hạng k tương đối nhỏ, chẳng hạn không lớn hơn 10, hoặc vài chục. Trong trường hợp này, biểu diễn affine cực đại (2) rất ngắn gọn, tức là số tham số cần thiết để mô tả f (tức là p) nhỏ hơn nhiều so với số tham số trong tập hợp dữ liệu ban đầu (tức là m(n+1)). Tuy nhiên, các phương pháp mà chúng ta mô tả không đòi hỏi k nhỏ.

Khi $\mathcal{F} = \mathcal{F}_{ma}^{k}$, bài toán khớp (1) rút về bài toán bình phương tối thiểu phi tuyến

Cực tiểu hóa…………………….

Với các biến $a_1, \dots\dots, a_k \epsilon \boldsymbol{R}^n$, $b_1, \dots\dots, b_k \epsilon \boldsymbol{R}$. Hàm J là hàm bậc hai từng đoạn của α. Thực vậy, đối với mỗi i, $f(u_i) - y_i$ tuyến tính từng đoạn, và J là tổng bình phương của các hàm này, vì vậy J là bậc hai lồi trên các vùng (đa diện) mà $f(u_i)$ affine. Nhưng J không lồi toàn bộ, vì vậy bài toán khớp (3) không lồi.

1.2 Tham số hóa tổng quát hơn

Chúng ta cũng sẽ xét một dạng được tham số hóa tổng quát hơn cho các hàm tuyến tính từng đoạn lồi,

…………………………………………..(4)

ở đây $\psi: \boldsymbol{R}^q \to \boldsymbol{R}$ " là hàm tuyến tính từng đoạn lồi (không đổi), và $\phi: \boldsymbol{R}^n \times \boldsymbol{R}^p \to \boldsymbol{R}^q$ là hàm song affine (không đổi). (Điều này có nghĩa là đối với mỗi x, $\phi(x, \alpha)$ là hàm affine của α và đối với mỗi α, $\phi(x, \alpha)$ là hàm affine của x.) Tham số hóa affine cực đại đơn giản (2) có dạng này, với $q=k$, $\psi(z_1, \dots\dots, z_k) = max(z_1, \dots\dots, z_k)$, và $\phi_i(x, \alpha) = a_i^T x + b_i$

Ví dụ, xét tập hợp hàm $\mathcal{F}$ là tổng của k số hạng, mỗi số hạng là cực đại của hai hàm affine,

………………………………………………(5)

Được tham số hóa bởi $a_1, \dots\dots, a_k, c_1, \dots\dots, c_k \epsilon \boldsymbol{R}^n$ và $b_1, \dots\dots, b_k, d_1, \dots\dots, d_k \epsilon \boldsymbol{R}$. Tập hợp này tương ứng với dạng tổng quát (4)

………………………………………………..

Và

…………………………………………….

thể lớn hơn rất nhiều biểu diễn dạng tổng quát ban đầu. Chẳng hạn như dạng hàm (5) đòi hỏi $p = 2k(n+1)$ tham số. Nếu hàm tương tự được viết dưới dạng affine cực đại, nó đòi hỏi $2k$ số hạng, và $2k(n+1)$ tham số. Hi vọng là dạng tổng quát được chọn tốt có thể cho chúng ta khớp cô đọng hơn của dữ liệu cho trước hơn so với dạng affine cực đại với cùng số tham số.

Xét một ví dụ khác về dạng tổng quát (4), xét trường hợp trong đó f được cho như giá trị tối ưu của quy hoạch tuyến tính (LP) với vế phải của phương trình phụ thuộc song affine vào x và các tham số:

…………………………………………………………………….

Ở đây c và A không đổi; b và B được xem là các tham số xác định f. Hàm này có thể được đặt ở dạng tổng quát (4) dùng

……………………………………………………………

Hàm ψ lồi và tuyến tính từng đoạn (chẳng hạn, xem, Boyd và Vandenberghe 2004); hàm ϕ tất nhiên là song affine theo x và *(b, B).*

1.3 Phép biến đổi và chuẩn hóa biến phụ thuộc

Chúng ta có thể áp dụng biến đổi affin không kì dị cho biến phụ thuộc u, bằng cách đặt

………………………………………………..

ở đây $T \epsilon R^{n \times n}$ không kì dị và $s \epsilon R^n$. Định nghĩa $\tilde{f}(\tilde{x}) = f\left(T^{-1}(x-s)\right)$, chúng ta có $\tilde{f}(\tilde{u}_i) = f(u_i)$. Nếu f tuyến tính từng đoạn và lồi, thì $\tilde{f}$ cũng vậy (và ngược lại). Miễn là $\mathcal{F}$ bất biến khi tổ hợp với các hàm affine, bài toán khớp dữ liệu (u_i, y_i) với hàm $f \epsilon \mathcal{F}$ giống như bài toán khớp dữ liệu $(\tilde{u}_i, y_i)$ với hàm $\tilde{f} \epsilon \mathcal{F}$.

Điều này cho phép chúng ta chuẩn hóa các dữ liệu biến phụ thuộc theo những cách khác nhau. Ví dụ, chúng ta có thể giả sử rằng nó có trung bình bằng không (mẫu) và hiệp phương sai bằng một (mẫu)

……………………………………………………(6)

Miễn là các dữ liệu u_i độc lập affine. (Nếu không, chúng ta có thể đưa bài toán về bái toán tương đương với chiều nhỏ hơn.)

1.4 Một số nét chung

Trong phần 2, chúng ta mô tả vài ứng dụng của khớp tuyến tính từng đoạn lồi. Trong phần 3, chúng ta mô tả giải thuật tìm tòi (gần đúng) để giải bài toán khớp affine cực đại (1). Giải thuật cơ bản này có vài nhược điểm, chẳng hạn như sự hội tụ về cực tiểu cục bộ xấu, hoặc không hội tụ gì cả. Bằng cách chạy giải thuật này một số lần hạn chế, từ các điểm ban đầu khác nhau, chúng ta thu được giải thuật khá tin cậy để khớp bình phương tối thiểu hàm affine cực đại của dữ liệu cho trước. Cuối cùng, chúng ta sẽ chỉ ra giải thuật này có thể được mở rộng để áp dụng cho trường hợp tham số hóa hàm tổng quát hơn (4) như thế nào. Trong phần 4 chúng ta đưa ra một số ví dụ bằng số.

1.5 Các công trình trước đây

Hàm tuyến tính từng đạo xuất hiện trong nhiều lĩnh vực và bối cảnh. Một số dạng tổng quát để biểu diễn hàm tuyến tính từng đoạn có thể được tìm thấy trong các tài liệu, chẳng hạn như, Kang and Chua, Kahlert và Chua (1978 , 1990). Vài phương pháo đã được đề xuất để khớp các hàm tuyến tính từng đoạn tổng quát của dữ liệu (nhiều chiều). Giải thuật mạng nơron được sử dụng trong Gothoskar và các cộng sự (2002); Phương pháp Gauss-Newton được sử dụng trong Julian và các cộng sự, Horst và Beichel (1998, 1997) để tìm gần đúng tuyến tính từng đoạn của các hàm trơn. Một nghiên cứu gần đây về các phương pháp bình phương tối thiểu với các hàm bán trơn là nghiên cứu của Kanzow and Petra (2004). Một quy trình lặp, ý tưởng tương tự như phương pháp của chúng ta, được mô tả trong Ferrari-Trecate và Muselli (2002). Phần mềm để khới các hàm tuyến tính từng đoạn tổng quát của dữ liệu được đề cập đến trong Torrisi and Bemporad (2004), Storace và De Feo (2002).

Trường hợp đặc biệt n =1, tức là, khớp một hàm trên R, bởi hàm tuyến tính từng đoạn đã được nghiên cứu rộng rãi. Ví dụ, một phương pháp để tìm số đoạn ít nhất để đạt được sai số cực đại cho trước được mô tả trong công trình của Dunham (1986); bài toán tương tự có thể đuợc tiếp cận dùng quy hoạch động (Goodrich 1994; Bellman và Roth 1969;Hakimiand Schmeichel 1991;Wanget al. 1993), hoặc giải thuật di truyền (Pittman và Murthy 2000). Bài toán đơn giản hóa hàm tuyến tính từng đoạn cho trước trên R, của hàm bao gồm số ít đoạn, được xét trong Imai

và Iri (1986).

Một bài toán có liên quan khác cũng nhận được nhiều sự chú ý là bài toán khớp đường cong tuyến tính từng đoạn, hoặc đa giác, trong R^2 của dữ liệu cho trước; chẳng hạn, xem Aggarwal và các cộng sự (1985), Mitchell va Suri (1992). Một thủ tụ lặp, có liên hệ chặt chẽ với giải thuật k-means và do đó về ý tưởng là giống với phương pháp của chúng ta, được mô tả trong công tình của Phillips và Rosenfeld (1988), Yin (1998).

Các hàm tuyến tính từng đoạn và các phép gần đúng đã được sử dụng trong nhiều ứng dụng, chẳng hạn như phát hiện các hoa văn trong ảnh (Rives và các cộng sự, 1985), đánh dấu đường biên dạng (Dobkin và các cộng sự, 1990), tách ra các đường thẳng trong các ảnh không gian (Venkateswar và Chellappa 1992), tối ưu hóa toàn cục (Mangasarian và các cộng sự, 2005), nén dữ liệu quy trình hóa học (Bakshi và Stephanopoulos 1996), và mô hình hóa mạhc (Ju-lian và các cộng sự;ChuaandDeng 1986;Vandenbergheetal. 1989).

Theo chúng tôi được biết thì chỉ có hai bài báo xét bài toán khớp hàm lồi tuyến tính từng đoạn của dữ liệu cho trước. Mangasarian và các cộng sự (2005) đã mô tả phương pháp tìm tòi để khớp hàm lồi tuyến tính từng đoạn có dạng $a + b^T x + \|Ax + c\|_1$ của dữ liệu cho trước (cùng với ràng buộc mà hàm đánh giá thấp dữ liệu). Mục tiêu chính của bài báo của họ là tìm underestimator lồi tuyến tính từng đoạn cho các hàm đã biết (không lồi), để sử dụng trong tối ưu hóa toàn cực; trái lại, mục tiêu của cúng ta chỉ đơn giản là khớp một số dữ liệu cho trước. Công trình có liên quan chayjt chẽ nhất mà chúng tôi biết là của Kim và các cộng sự (2004). Trong bài báo này, Kim và các cộng sự đã mô tả phương pháp khớp hàm affine cực đại (lồi) của dữ liệu cho trước, tăng số lượng số hạn để có được khớp tốt hơn. (Quả thực, chúng tôi mô tả phương pháp khớp hàm đơn thức cực đại của các mô hình mạch; xem Phần. 2.3 .)

2 Các ứng dụng

Trong phần này, chúng tôi mô tả vắn tắt một số ứng dụng của khớp tuyến tính từng đoạn lồi. Phần này không được sử dụng lại ở các mục tiếp theo.

2.1 Mô hình hóa LP

Một ứng dụng là trong mô hình hóa LP, tức là, xây dựng công thức gần đúng của một bài toán thực tế như một LP. Giả sử một bài toán được mô hình hóa khá tốt dùng các ràng buộc đẳng thức và bất đẳng thức tuyến tính, với một số ít ràng buộc bất đẳng thức phi tuyến. Bằng cách lấy gần đúng những hàm phi tuyến này là các hàm tuyến tính từng đoạn lồi, toàn bộ bài toán có thể được phát biểu như một LP

, và do đó có thể được giải tốt.

Chẳng hạn như, xét bài toán điều khiển tối ưu nhiên liệu cực tiểu, với hàm tiêu hao nhiên liệu động học và phi tuyến,

……………………………………………..

Với các biến $x(0), \ldots \ldots x(T) \epsilon R^n$(quỹ đạo trạng thái), và $u(0), \ldots \ldots, u(T-1) \epsilon R^m$ (đầu vào điều khiển). Dữ liệu bài toán là $A(0), \ldots.., A(T-1)$ (ma trận động học), $B(0), \ldots.., B(T-1)$ (các ma trận điều khiển), x_{int} (trạng thái ban đầu), và x_{des} (trạng thái quan tâm cuối cùng). Hàm $f: R^m \rightarrow R$ là hàm tiêu hao nhiên liệu, nó cho chúng ta biết nhiên liệu được sử dụng trong một chu kì, như hàm theo giá trị đầu vào điều khiển. Bây giờ giả sử chúng ta có các dữ liệu kinh nghiệm hoặc đo được của một số giá trị của đầu vào điều khiển $u \epsilon R^m$, cùng với hàm sử dụng nhiên liệu $f(u)$. Nếu chúng ta có thể khớp những dữ liệu này với hàm tuyến tính từng đoạn lồi, ví dụ,

…………………………………………………………………..

Thì chúng ta có thể phát biểu (gần đúng) bài toán điều khiển tối ưu nhiên liệu cực tiểu là

Cực tiểu…………………….

Thõa mãn điều kiện…………………………

Với các biến………………….

2.2 Đơn giản hóa các hàm lồi

Một ứng dụng khác của khớp tuyến tính từng đoạn lồi là đơn giản hóa các hàm lồi phức tạp, hoặc cần tính toán. Để minh họa ý tưởng này, chúng ta tiếp tục bài toán điều khiển tối ưu nhiên liệu cực tiểu được mô tả ở trên, với hàm sử dụng nhiên liệu tuyến tính từng đoạn. Xét hàm V : $R^n \rightarrow R$, nó ánh xạ trạng thái ban đầu x_{int} vào hàm sử dụng nhiên liệu cực tiểu tương ứng của nó, tức là, giá trị tối ưu của LP (7). (Đây là hàm giá trị Bellman đối với bài toán điều khiển tối ưu.) Hàm giá trị là tuyến tính từng đoạn và lồi, nhưng có thể đòi hỏi một số lượng lớn các số hạng được biểu diễn dạng affine cực đại. Chúng ta có thể (chỉ là khả năng) hình thành nên một gần đúng đơn giản của V qua một hàm affine cực đại với ít số hạng hơn như sau. Trước hết, chúng ta ước tính V qua LP (7), cho một số lượng lớn các điều kiện ban đầu. Sau đó, chúng ta khớp hàm affine cực đại với một số lượng hạn chế các số hạng của dữ liệu cuối cùng. Hàm giá trị gần đúng tuyến tính từng đoạn lồi này có thể được sử dụng để xây dựng bộ điều khiển phản hồi đơn giản cực tiểu hóa việc sử dụng nhiên liệu; chẳng hạn, xem Bemporad và các cộng sự (2002).

2.3 Khớp đơn thức cực đại cho quy hoạch hình học

Khớp affine cực đại có thể được dùng để tìm gần đúng đơn thức cực đại của một hàm dương, để sử dụng trong mô hình hóa quy hoạch hình học; xem Boyd và các cộng sự (2006). Từ dữ liệu cho trước $\left(z_i, w_i \epsilon R^n_{++} \times R_{++}\right)$, chúng ta đặt

……………………………………………………………

(log của một vector được biểu diễn theo từng thành phần.) Bây giờ chúng ta khớp dữ liệu này với một mô hình affine cực đại,

…………………………………………………

Điều này cho chúng ta một mô hình đơn thức cực đại

……………………………………………………..

ở đây g_i là các hàm đơn thức

……………………………………….

(Những hàm này không đơn thức theo nghĩa thông thường mà đơn thức theo nghĩa được dùng trong quy hoạch hình học.)

3Giả thuật phân vùng bình phương tối thiểu

3.1 Giải thuật

Trong phần này chúng ta đưa vào giải thuật tìm tòi để giải (gần đúng) bài toán khớp affine cực đại số hạng k(3), tức là,

Cực tiểu……………………………………….

Với các biến $a_1, \ldots\ldots, a_k \epsilon R^n$ và $b_1, \ldots\ldots, b_k \epsilon R$. Giải thuật luân phiên giữa hai thao tác phân vùng dữ liệu và khớp bình phương tối thiêu để cập nhật các hệ số.

Chúng ta đặt $P_j^{(l)}$ với j=1,…,k, là phân cùng của chỉ số dữ liệu ở lần lặp thứ l,

Tức là, $P_j^{(l)} \subseteq \{1, \ldots, m\}$, với

…………………………………………………

(Chúng ta sẽ mô tả các phương pháp chọn phân vùng ban đầu $P_j^{(0)}$ sau)

Đặt $a_j^{(l)}$ và $b_j^{(l)}$ chỉ giá trị của các tham số ở lần lặp thứ l của giải thuật. Chúng ta tạo ra các giá trị tiếp theo $a_j^{(l+1)}$ và $b_j^{(l+1)}$, từ phân vùng hiện tại $P_j^{(l)}$ như sau. Đối với mỗi $j = 1, \ldots, k$, chúng ta thực hiện khớp bình phương tối thiểu $a_j^T u_i + b_j$ của y_i, chỉ dùng các điểm dữ liệu với $i \in P_j^{(l)}$. Nói cách khác, chúng ta chọn $a_j^{(l+1)}$ và $b_j^{(l+1)}$ như các giá trị của a và b cực tiểu hóa

…………………………………………….(8)

Trong trường hợp đơn giản nhất (và phổ biến nhất), có một cặp (a, b) duy nhất cực tiểu hóa (8), tức là,

…………………………………………………………..(9)

ở đây tổng được lấy trên $i \in P_j^{(l)}$.

Khi có nhiều cực tiểu của hàm bậc hai (8), tức là ma trận nghịch đảo của (9) kì dị, chúng ta có vài tùy chọn. Một tùy chọn là cộng một số chính quy hóa nào đó vào đối tượng bình phương tối thiểu đơn giản trong (8), tức là, số hạng thêm vào có dạng $\lambda\|a\|_2^2 + \mu b^2$, ở đây λ và μ là các hằng số dương. Một khả năng khác là chọn các tham số được cập nhật như cực tiểu duy nhất của (8) gần với giá trị trước nhất $\left(a_j^{(l)}, b_j^{(l)}\right)$, trong chuẩn Euclide.

Dùng các giá trị mới của các hệ số, chúng ta cập nhật phân vùng để thu được $P_j^{(l+1)}$, bằng cách quy i cho $P_s^{(l+1)}$ nếu

……………………………………….(10)

(Điều này có nghĩa là số hạng $a_j^{(l)T}u_i + b_j^{(l)}$ là “tích cực” tại điểm dữ liệu u_i.) Nói một cách gần đúng điều này có nghĩa là $P_j^{(l+1)}$ là tập hợp các chỉ số mà hàm affin $a_j^T z + b_j$đạt cực đại với nó; chúng ta có thể phá vỡ các ràng buộc (nếu có) tùy ý.

Phép lặp này chạy cho đến khi hội tụ, nó xuất hiện nếu phân vùng tại một lần lặp giống như phân vùng tại lần lặp trước, hoặc đạt được một số lần lặp cực đại.

Chúng ta có thể viết giải thuật là

GIẢI THUẬT PHÂN VÙNG BÌNH PHƯƠNG TỐI THIỂU

Phân vùng cho trước $P_1^{(0)}, \ldots., P_K^{(0)}$ của $\{1, \ldots, m\}$, giới hạn lặp l_{max}

Đối với $l = 0, \ldots, l_{max}$

1. Tính $a_j^{(l+1)}$ và $a_j^{(l+1)}$ như trong (9)
2. Hình thành phân vùng $P_1^{(l+1)}, \ldots \ldots, P_k^{(l+1)}$ như trong (10)
3. Thoát nếu $P_j^{(l)} = P_j^{(l+1)}$ đối với $j = 1, \ldots, k$

Trong suốt quá trình thực thi giải thuật phân vùng bình phương tối thiểu, một hoặc nhiều tập hợp $P_j^{(l)}$ có thể rỗng. Cách đơn giản nhất là bỏ các tập hợp rỗng từ phân vùng, và tiếp tục với các giá trị k nhỏ hơn.

3.2 Diễn giải dưới dạng phương pháp Gauss-Newton

Chúng ta có thể diễn giải giải thuật của bài toán (3) dưới dạng phương pháp Gauss - Newton. Giả sử rằng một điểm $u \epsilon R^n$, có một j duy nhất mà đối với nó $a_j^T u + b_j$ (tức là không có ràng buộc nào đối với cực đại định nghĩa $f(u)$. Trong trường hợp này, f khả vi đối với a và b; thực sự, nó là affin cục bộ trong các giá trị tham số này. Gần đúng bậc nhất của nó tại a và b là

……………………………………..

Phép gần đúng này sẽ chính xác nếu các giá trị tham số nhiễu $\tilde{a}_1, \ldots.., \tilde{a}_k$, $\tilde{b}_1, \ldots.., \tilde{a}_b$ đủ gần với các giá trị tham số $a_1, \ldots., a_k, b_1, \ldots., a_b$.

Bây giờ giả sử rằng mỗi điểm dữ liệu u_i mà đối với nó $f(u_i) = a_j^{(l)T} u_i + b_j^{(l)}$ (tức là không có ràng buộc nào lên cực đại xác định $f(u_i)$. Thế thì phép gần đúng bậc nhất của $\left(f(u_1), \dots\dots, f(u_m)\right)$ là

…………………………..

ở đây $j(i)$ là j tích cực duy nhất tại u_i, tức là $i \epsilon P_j^{(l)}$.

Trong phương pháp Gauss-Newton đối với bài toán bình phương tối thiểu phi tuyến, chúng ta hình thành nên gần đúng bậc nhất của đối số của chuẩn, và giải bài toán bình phương tối thiểu cuối cùng để nhận được bước lặp tiếp theo. Thế thì, trong trường hợp này, chúng ta hình thành nên bài toán cực tiểu hóa bình phương tối thiểu tuyến tính

…………………………………………….

Trên các biến $\tilde{a}_1, \dots\dots, \tilde{a}_k$, $\tilde{b}_1, \dots\dots, \tilde{b}_k$. Chúng ta có thể sắp xếp lại tổng xác định J thành các số hạng liên quan đến các cặp biến $a_1, b_1, \dots\dots a_k, b_k$ riêng biệt:

……………………………………………….

Ở đây

…………………………………………………….

Tất nhiên, chúng ta có thể cực tiểu hóa $\hat{J}$ bằng cách cực tiểu hóa riêng biệt từng $\hat{J}_i$. Hơn nữa, các giá trị tham số cực tiểu hóa $\hat{J}$ đúng bằng $a_1^{(l+1)}, \dots\dots\dots, a_k^{(l+1)}, b_1^{(l+1)}, \dots\dots\dots, b_k^{(l+1)}$. Đây đúng là giải thuật phân vùng bình phương tối thiểu được mô tả ở trên.

Giải thuật này có liên quan mật thiết với giải thuật k-means được sử dụng trong gom cụm bình phương tối thiểu (Gersho và Gray 1991). Giải thuật k-means giải bài toán tìm tập hợp các điểm k trong R^n, {z_1 ,...,z_k}, cực tiểu hóa khoảng cách Euclide bình phương trung bình của tập hợp các dữ liệu cho trước $u_1, \dots, u_m \epsilon R^n$. (Khoảng cách giữa một điểm u và tập hợp các điểm {z_1 ,...,z_k} được định nghĩa là khoảng cách cực tiểu, tức là $min_{j=1,\dots,k} \|u - z_i\|_2$. Trong giải thuật k-means, chúng ta lặp lại giữa hai bước:

Thứ nhất, chúng ta phân vùng các điểm dữ liệu theo điểm hiện tại gần nhất trong tập hợp $\{z_1 ,...,z_k\}$; sau đó chúng ta cập nhật mỗi z_i như trung bình của các điểm trong phân cùng có liên quan. (Trung bình cực tiểu hóa tổng của các bình phương của khoảng cách Euclide của điểm). Về mặt khái niệm, giải thuật của chúng ta giống với giải thuật k-means: chúng ta phân vùng các điểm dữ liệu mà hàm affine tích cực với nó (tức là lớn nhất) và sau đó cập nhật các hàm affine một cách riêng biệt, chỉ dùng các điểm dữ liệu trong phân vùng liên quan của nó.

3.3 Sự không phân kì của giải thuật phân vùng bình phương tối thiểu

Giải thuật phân vùng bình phương tối thiểu cơ bản không cần hội tụ; nó có thể đưa vào một chu trình giới hạn (thay đổi). Chẳng hạn xét dữ liệu

………………………………..

và k = 2. Hiển nhiên, dữ liệu kiểu này không thể được khớp tốt bằng bất kì hàm lồi nào; khớp tốt nhấ (toàn cục) có thể thu được bằng hàm hằng $f(u) = 1$. Tuy nhiên, đối với nhiều giá trị tham số ban đầu, giải thuật hội tụ về chu trình giới hạn với chu kì bằng 2, tuần tự giữa hai hàm

……………………………………….

Do đó, giải thuật không hội tụ; hơn nữa, mỗi hàm f_1 và f_2 cho chúng ta sự khớp gần tối ưu của các dữ liệu.

Mặt khác, với dữ liệu thực tế (không được thiết kế để minh họa sự không hội tụ), chúng ta thấy rằng giải thuật phân vùng bình phương tối thiểu hội tụ trong đa số trường hợp. Trong bất cứ trường hợp nào, sự không hội tụ không có những hệ quả thực tế bởi vì giải thuật kết thúc sao một số bước cực đại xác định, và hơn nữa, chúng tôi đề nghị rằng có thể chạy từ một số điểm bắt đầu, với khớp tốt nhất được dùng như khớp cuối cùng.

3.4 Giải thuật khớp tuyến tính từng đoạn

Giải thuật phân vùng bình phương tối thiểu, được sử dụng bởi chính nó, có vài hạn chế nhất định. Nó không cần hội tụ, và khi nó hội tụ, nó có thể (và thường) hội tụ đến gần đúng tuyến tính từng đoạn với sự khớp dữ liệu không tốt. Cả hai vấn đề này có thể được giảm nhẹ bằng cách chạy giải thuật phân vùng bình phương tối thiểu nhiều lần, với các phân vùng ban đầu khác nhau. Khớp cuối cùng được chọn là khớp tốt nhất thu được trong số tất cả những lần lặp của những lần chạy giải thuật.

Trước hết chúng ta mô tả phương pháp đơn giản để tạo phân vùng ban đầu ngẫu nhiên. Chúng ta chọn ngẫu nhiên các điểm $p_1, \ldots \ldots . . , p_K$, và định nghĩa phân vùng ban đầu là các tập hợp Voronoi gắn với những điểm này. Chúng ta có

…………………………………………………………(11)

(Vì thế, $P_j^{(0)}$là tập hợp các chỉ số của các điểm dữ liệu gần p_j nhất.) Các điểm nhân p_i sẽ được tạo ra theo phân bố nào đó khớp với hình dạng của các điểm dữ liệu u_i

, chẳng hạn, chúng có thể được chọn từ phân bố bình thường với giá trị trung bình

$\bar{u}$ và hiệp phương sai $\sum_u$ (xem (6)).

Toàn bộ giải thuật có thể được mô tả là

GIẢI THUẬT KHỚP TUYẾN TÍNH TỪNG ĐOẠN

Cho trước số lần thử N_{trial} , giới hạn lặp l_{max}

Đối với $i = 1, \ldots . . , N_{trials}$

1. Tạo ngẫu nhiên phân vùng ban đầu qua (11).

2. Chạy giải thuật phân vùng bình phương tối thiểu với giới hạn lặp l_{max} .

3. Theo dõi khớp RMS tốt nhất thu được.

3.5 Khớp dạng tổng quát

Trong phần này, chúng ta sẽ chỉ ra giải thuật phân vùng bình phương tối thiểu có thể được điều chỉnh để khớp các hàm tuyến tính từng đoạn với dạng tổng quát hơn (4),

……………………………………………..

ở đây ψ là hàm tuyến tính từng đoạn lồi không đổi, và ϕ là hàm song affine.

Chúng ta mô tả giải thuật phân vùng bình phương tối thiểu theo phân vùng của các chỉ số, mà theo đó k hàm affine tích cực tại điểm u_i. Cách tiếp cận tương tự cho một phân vùng tường minh sẽ khống đúng trong trường hợp tổng quát, bởi vì kích thướt của phân vùng có thể cực kì lớn. Thay vì vậy, chúng ta bắt đầu từ ý tưởng là phân vùng cho gần đúng của $f(u_i)$ là affine theo α, và đúng gần $\alpha^{(l)}$. Nếu có ràng buộc tại u_i (tức là có một hàm affine duy nhất đạt cực đại), thì gần đúng affine chính xác tại lân cận của giá trị tham số hiện tại $\alpha^{(l)}$.

Chúng ta có thể làm điều tương tự với dạng tổng quát hơn. Đối với mỗi I, chúng ta tìm $a_i(\alpha)$ và $b_i(\alpha)$, cả hai hàm affine của α, sao cho

………………………………………………..

Đối với α gần $\alpha^{(l)}$, giá trị hiện tại của các tham số. Gần đúng này là chính xác trong lân cận của $\alpha^{(l)}$ nếu $\psi(u_i, \alpha)$ là điểm khả vi của ψ. (Đối với hàm affine cực đại, đây là trường hợp khi không có "ràng buộc" trên u_i. Nếu nó không phải là một điểm như thế, chúng ta có thể chọn bất cứ mô hình gradient dưới nào của $f(u_i, \alpha)$, tức là bất cứ $a_i(\alpha)$ và $b_i(\alpha)$ nào sao cho

……………………………………..

(phép gần đúng này chính xác khi $\alpha = \alpha^{(l)}$), và

………………………………………………..

Đối với mọi α (Trong trường hợp hàm affine cực đại, phá vỡ bất kì ràng buộc nào cũng thỏa mãn điều kiện này)

Sau đó chúng ta tính các giá trị tham số mới dùng phương pháp dạng Gauss - Newton. Chúng ta thay $f(u_i)$ trong biểu thức cho J với

…………………………………………..

Nó affine theo α. Sau đó, chúng ta chọn $\alpha^{(l+1)}$ là cực tiểu của

…………………………………………

Có thể tìm được dùng bình phương tối thiểu tuyến tính tiêu chuẩn.

Để giảm quy tắc cập nhật này, chúng ta có thể cộng số hạng chính quy hóa vào $\hat{J}$ bằng cách chọn $\alpha^{(l+1)}$ là cực tiểu của

……………………………………

4 Các ví dụ bằng số

Trong phần này chúng ta trình bày một số kết quả bằng số cụ thể, dùng tập hợp dữ liệu sau. Số chiều là n = 3, và chúng ta có m = 11 3 = 1331 điểm. Tập hợp các điểm u_i là $\nu \times \nu \times \nu$, ở đây $\nu = \{-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5\}$. Các giá trị thu được là $y_i = g(u_i)$, ở đây

,where g là hàm (lồi)

……………………………………

Chúng ta sử dựng giải thuật khớp tuyến tính từng đoạn được mô tả trong phần 3.4 ,với giới hạn lặp l_{max} = 50 , và số số hạng thay đổi từ k = 0 đến k = 20 . Đối với k = 0, hàm khớp được chọn bằng không, vì vậy chúng ta báo cáo giá trị RMS của $y_1, \ldots\ldots, y_m$ là khớp RMS. Đối với k=1, khớp là khớp affine tốt nhất của dữ liệu, có thể tìm được dùng bình phương tối thiểu. Hình 1 biểu diễn khớp RMS thu được sau N_{trial} = 10 lần thử (đường cong ở trên), và sau N_{trial} = 100 lần thử (đường cong bên dưới). Những đường cong này biểu diễn khớp tốt

…………………………….

Thu được chỉ với 10 lần thử, và những cái tốt hơn thu được với 100 lần thử.

Để có ý tưởng tốt hơn về sự thay đổi khớp RMS thu được với các lần thử khác nhau, cũng như số bước cần thiết để hội tụ (nếu có), chúng ta cố định số số hạng

k = 12 , và chạy giải thuật phân vùng bình phương tối thiểu 200 lần, với giới hạn 50 lần lặp, ghi nhận cả hai khớp RMS thu được, và số bước trước khi hội tụ. (Số bước được báo cáo là 50 nếu giải thuật phân vùng bình phương tối thiểu không hội tụ trong 50 bước.) Hình 2 biểu diễnđồ thị của khớp RMS thu được. Chúng ta có thể thấy rằng khớp thường, nhưng không luôn luôn, khá tốt; chỉ trong vài trường hợp, khớp thu được xấu. Hiển nhiên, cái tốt nhất của một số ít lần thử sẽ khá tốt.

Hình 3 biểu diễn phân bố của số lần lặp của giải thuật phân vùng bình phương tối thiểu cần thiết để hội tụ. Hội tụ không xảy ra ở 13 trong số 200 lần thử. Nhưng quả thực, khớp RMS thu được trong những lần thử này không tồi. Thông thường, sự hội tụ xuất hiện trong khoảng 25 lần lặp.

Ví dụ bằng số cuối cùng của chúng ta, chúng ta so sánh khớp dữ liệu với hàm affine cực đại với k số hạng, và với dạng tổng quát hơn

……………………………….

Được tham số hóa bởi $a_1, \dots\dots, a_k \epsilon R^n$ và $b_1, \dots\dots, b_k$. (Chú ý rằng số tham số trong mỗi hàm giống nhau.) Hàm này tương ứng với dạng tổng quát (4) với

……………………………….(Hình vẽ)

……………………..(công thức)

Và

………………………………

Chúng ta thiết lập giới hạn lặp đối với cả hai dạng l_{max} = 100 , và chọn khớp tốt nhất thu được trong N_{trial}=10 lần thử. Chúng ta dùng giá trị $\rho = 10^{-5}$ đối với tham số chính quy hóa trong giải thuật dạng tổng quát. Hình 4 biểu diễn khớp RMS thu được đối với cả hai dạng, theo k. Tất nhiên, dạng tổng-cực đại cho khớp RMS hơi tốt hơn dạng affine cực đại.

5 Kết luận

Chúng tôi đã mô tả một phương pháp mới để khớp hàm tuyến tính từng đoạn lồi của một tập hợp dữ liệu cho trước (có thể rất lớn) (với một số biến độc lập vừa phải). Phương pháp này theo kiểu tìm tòi, bởi vì giải thuật có thể không hội tụ về khớp tối ưu toàn cục. Tuy nhiên, các ví dụ bằng số cho thấy rằng phương pháp có thể áp dụng tốt trong thực tế trên dữ liệu có thể được khớp tốt bằng hàm lồi.

Phương pháp này có nhiều ứng dụng trong mô hình hóa tối ư thực tế. Các mẫu dữ liệu cũng có thể được sử dụng để tạo ra các hàm lồi tuyến tính từng đoạn, và nó có thể được sử dụng để xây dựng các mô hình lập trình tuyến tính.