GIẢI THUẬT SMO CHO BÀI TOÁN SVM

Biên soạn: Phạm Nguyên Khang

1 Giải thuật SMO

Giải thuật SMO (sequential minimal optimization: cực tiểu hoá tuần tự) do John Platt đề xuất là một phương pháp hiệu quả để giải bài toán đối ngẫu của SVM.

1.1 Phương pháp tăng toạ độ (Coordinate ascent)

Xét bài toán tối ưu không ràng buộc sau:

$$\max_{\alpha} W(\alpha_1, \alpha_2, \dots, \alpha_m)$$

Ta có thể xem W như một hàm của các tham số α. Ngoài 2 giải thuật tối ưu là *giảm gradient* và *phương pháp Newton*, ta có thể sử dụng giải thuật *tăng toạ độ* (coordinate ascent) để tối ưu W.

Loop until convergence {

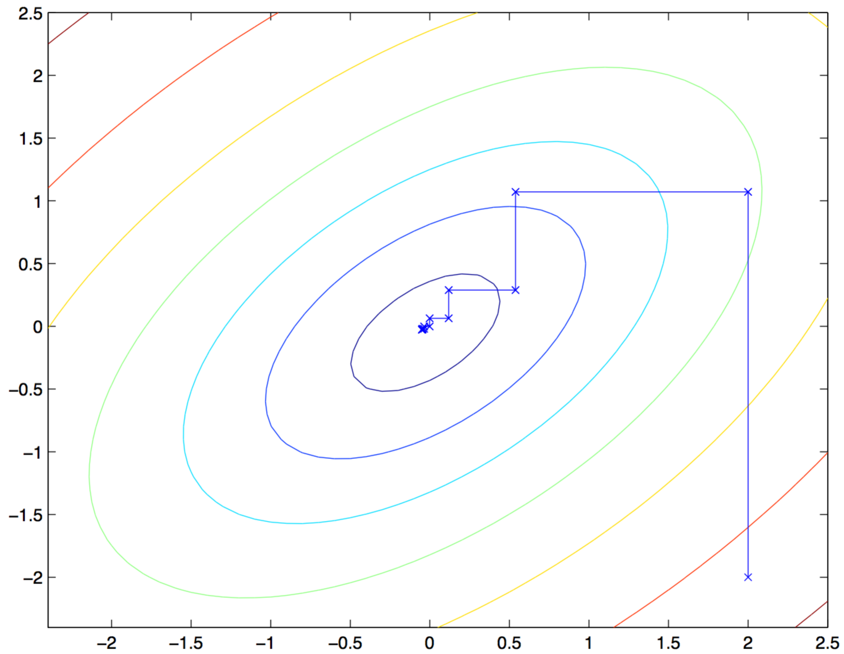
for $i = 1, 2, \dots, m$ {

$$\alpha_i = arg \max_{\hat{\alpha}_i} W(\alpha_1, \alpha_2, \dots, \hat{\alpha}_i, \dots, \alpha_m)$$

}

Trong vòng lặp for của giải thuật, ta cố định tất cả các biến trừ α_i, và tối ưu W theo α_i. Trong giải thuật này, thứ tự của các tham số không quan trọng. Ta có thể chọn α_i để tối ưu sao cho W được tăng nhiều nhất có thể.

Khi bài toán "arg max" trong vòng lặp for có thể giải dễ dàng và nhanh chóng thì giải thuật tăng toạ độ là một lựa chọn hiệu quả cho trường hợp này. Hình vẽ bên dưới minh hoạ nguyên lý hoạt động của giải thuật:



Các đường ellipse trong hình trên là đường biên của *hàm toàn phương* (quaratic function) mà ta muốn tối ưu. Giải thuật tăng toạ độ được khởi tạo tại điểm (2, -2). Đường zig-zag trong hình là đường đi từ điểm khởi tạo đến điểm tối ưu toàn cục. Chú ý là ở mỗi bước lặp, giải thuật tăng toạ độ di chuyển điểm tối ưu theo hướng song song với 1 trong 2 trục toạ độ vì mỗi lần ta chỉ tối ưu trên một biến.

1.2 Giải thuật SMO

Xét bài toán tối ưu đối ngẫu của SVM:

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle \tag{1}$$

với các ràng buộc:

$$0 \le \alpha_i \le C, i = 1, \dots, m \tag{2}$$

$$\sum_{i=1}^{m} \alpha_i y^{(i)} = 0 \tag{3}$$

Giả sử ta cố định các tham số $\alpha_2, \dots, \alpha_m$ và tối ưu W theo α_1. Ta có thể làm được không ? Câu trả lời là không vì theo ràng buộc (3), ta có:

$$\alpha_1 y^{(1)} = -\sum_{i=2}^{m} \alpha_i y^{(i)}.$$

Nhân 2 vế cho $y^{(1)}$ ta được:

$$\alpha_1 = -y^{(1)} \sum_{i=2}^{m} \alpha_i y^{(i)}.$$

(vì $y^{(1)} \in \{-1,1\}$ nên $y^{(1)} y^{(1)} = 1$.)

Do đó, ta luôn xác định được α_1 từ các $\alpha_2, \dots, \alpha_m$. Và như thế ta không thể cải thiện được W nếu cố định m – 1 biến.

Để cải thiện W, ta cần cập nhật ít nhất 2 biến α_i và α_j đồng thời. Điều này chính là mấu chốt của giải thuật SMO.

Loop until convergence {

1. Chọn 1 cặp α_i và α_j (có thể sử dụng heuristic để chọn 2 biến "tốt nhất" có thể cải thiện W nhiều nhất)
2. Tối ưu W theo α_i và α_j trong khi cố định các biến khác.

}

Để kiểm tra điều kiện hội tụ, ta có thể kiểm tra điều kiện KKT (với một dung sai tol nào đó, vd: 0.001).

Lý do chính mà SMO là một giải thuật hiệu quả là do bước 2 trong giải thuật (cập nhật α_i và α_j) có thể được thực hiện một cách nhanh chóng.

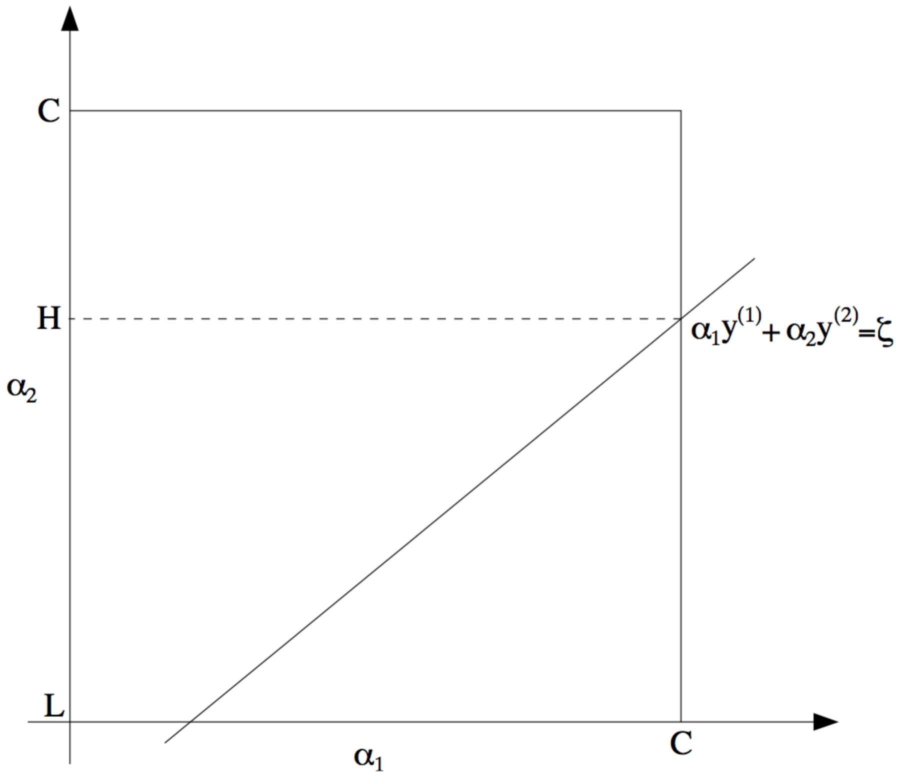
Giả sử ta cố định $\alpha_3, \ldots, \alpha_m$ và tối ưu W theo α_1 và α_2. Từ ràng buộc (3) ta có:

$$\alpha_1 y^{(1)} + \alpha_2 y^{(2)} = -\sum_{i=3}^{m} \alpha_i y^{(i)}.$$

Vế phải của phương trình này là hằng số và ta đặt nó bằng ζ (zeta):

$$\alpha_1 y^{(1)} + \alpha_2 y^{(2)} = \zeta. \tag{4}$$

Ta có thể minh hoạ ràng buộc (4) bằng hình vẽ bên dưới:



Từ ràng buộc (2), ta biết rằng α_1 và α_2 nằm trong hình vuông $[0, C] \times [0, C]$. Ràng buộc (4) là đường thẳng $\alpha_1 y^{(1)} + \alpha_2 y^{(2)} = \zeta$. Cũng từ các ràng buộc này ta biết rằng $L \leq \alpha_2 \leq H$. Nếu không (α_1, α_2) không thể đồng thời nằm trong hộp và nằm trên đường thẳng. Trong ví dụ này L = 0, tuy nhiên trong trường hợp tổng quát L và H là một khoảng giá trị có thể của α_2 để α_1 và α_2 nằm trong hình vuông $[0, C] \times [0, C]$.

Dùng công thức (4), ta có thể biểu diễn α_1 như một hàm của α_2:

$$\alpha_1 = \left(\zeta - \alpha_2 y^{(2)}\right) y^{(1)}.$$

Như vậy hàm mục tiêu $W(\alpha)$ có thể viết lại:

$$W(\alpha_1, \alpha_2, \ldots, \alpha_m) = W\left(\left(\zeta - \alpha_2 y^{(2)}\right) y^{(1)}, \alpha_2, \ldots, \alpha_m\right).$$

Xem $\alpha_3, \ldots, \alpha_m$ là hằng số, ta có thể xem W như một hàm bậc 2 theo α_2:

$$W(\alpha_2) = a\alpha_2^2 + b\alpha_2 + c \tag{5}$$

Nếu bỏ qua ràng buộc (2) (ràng buộc nằm trong hình vuông) hoặc bỏ qua ràng buộc $L \leq \alpha_2 \leq H$ thì ta có thể dễ dàng cực đại hoá W bằng cách lấy đạo hàm theo α_1 và giải phương trình đạo hàm = 0.

Giả sử $\alpha_2^{new,unclipped}$ là nghiệm của phương trình đạo hàm bằng 0. Ta có thể tìm α_2 thoả mãn ràng buộc (2) và (4) bằng cách "bắt" (clipping) nó nằm trong khoảng [L, H] hay:

$$\alpha_2^{new} = \begin{cases} H & \textit{nếu } \alpha_2^{new,unclipped} > H \\ \alpha_2^{new,unclipped} & \textit{nếu } L \leq \alpha_2^{new,unclipped} \leq H \\ L & \textit{nếu } \alpha_2^{new,unclipped} < L \end{cases}$$

Sau khi đã tìm được α_2^{new}, ta có thể sử dụng công thức (4) để tìm α_1^{new}.

Để tìm hiểu chi tiết hơn về giải thuật SMO, bạn đọc có thể tìm hiểu thêm trong bài báo của Platt ví dụ như: heuristic dùng để chọn α_i và α_j hoặc cập nhật giá trị b như thế nào khi chạy giải thuật SMO.

Bài tập: Hãy biểu diễn W trong công thức (1) về dạng hàm bậc 2 theo công thức (5), lấy đạo hàm theo α_2 và giải phương trình đạo hàm = 0 để tìm α_2.