1 Proven

1. Phụ đề (*)

   Ta có: Vector từ điểm $O(0,0)$ đến mặt phẳng $\vec{w}\vec{x}+c=0(d)$ hay vector từ điểm $O(0,0)$ đến hình chiếu của điểm O lên mặt phẳng d là:

   $$(\vec{O,d}) = \frac{-c}{||\vec{w}||}$$

   Từ đây ta có: Vector từ điểm I đến mặt phẳng $\vec{w}\vec{x}+c=0(d)$ là:

   $$(\vec{I,d}) = -(\vec{O,d'}) + (\vec{O,d})$$

   Với $\vec{w}(\vec{x}-\vec{I})=0(d')$ là mặt phẳng đi qua I và song song với mặt phẳng d

   $$\Rightarrow (\vec{I,d}) = -\frac{\vec{w}\vec{I}}{||\vec{w}||} + \frac{-c}{||\vec{w}||}$$

   $$\Rightarrow (\vec{I,d}) = -\frac{\vec{w}\vec{I}+c}{||\vec{w}||}$$

   Từ đó khoảng cách từ điểm I đến mặt phẳng $\vec{w}\vec{x}+c=0(d)$ là:

   $$d(d,O) = ||\frac{\vec{w}\vec{I}+c}{||\vec{w}||}||$$

2. Phụ đề (**)
   Xét bài toán: Tìm các điểm cực trị của hàm số $w=f(x,y)$ (1.1)
   Với điều kiện ràng buộc $g(x,y)=b$ (1.2)
   Trong đó x và y là các biến chọn, w là biến mục tiêu, f là hàm mục tiêu
   Phương pháp nhân tử Lagrange:
   Xuất phát từ hàm mục tiêu (1.1) và điều kiện (1.2) ta lập hàm số sau (gọi là hàm Lagrange):

   $$L(x,y,\lambda) = f(x,y) + \lambda[b-g(x,y)](1.3)$$

   Hàm số (1.3) có thêm một biến phụ λ, gọi là nhân tử Lagrange. Chú ý rằng với tất cả các điểm $M(x,y)$ thoản mãn điều kiện (1.2), tức là khi xét cặp biến chọn (x,y) trong miền biến thiên đã bị thu hẹp bởi điều kiện (1.2), hàm mục tiêu w đồng nhất với hàm số L. Định lý sau đây cho biết mối liên hệ giữa hàm số Lagrange (1.3) và bài toán cực trị có điều kiện mà ta đang xem xét.
   Định lý:
   Giả sử các hàm số $f(x,y)$ và $g(x,y)$ có đạo hàm riêng liên tục trong một lân cận của điểm $M_0(x_0,y_0)$ và $g'_y(x_0,y_0)\neq 0$. Khi đó nếu hàm số (1.1),

với điều kiện (1.2) đạt cực trị tại điểm $M_0(x_0, y_0)$ thì tồn tại bộ ba số thực $x_0, y_0, \lambda_0)$ là nghiệm của hệ phương trình:

$$\begin{cases} L'_\lambda = b - g(x,y) \\ L'x = f'_x - \lambda g'_x = 0 \\ L'y = f'_y - \lambda g'_y = 0 \end{cases}$$

Ý nghĩa của nhân tử Lagrange

Trong mô hình bài toán cực trị có điều kiện ta phân biệt vai trò của các biến số $x_1, x_2, \ldots x_n, w$ với tham số b. Các biến chọn $x_1, x_2, \ldots x_n$ và biến mục tiêu w được gọi là các biến nội sinh, do bản thân mô hình quyết định thông qua phương pháp nhân tử Lagrange. Khác với biến nội sinh, tham số b là một hằng số cho trước, không do mô hình quyết định. Phương án chọn tối ưu $\bar{X}(\bar{x_1}, \bar{x_2}, \ldots \bar{x_n})$ của bài toán và giá trị tối ưu $\bar{w}$ của hàm mục tiêu phụ thuộc vào b.

$\bar{x_1} = \bar{x_1}(b), \bar{x_2} = \bar{x_2}(b), \ldots, \bar{x_n} = \bar{x_n}(b)$

Theo phương pháp nhân tử Lagrange, phương án chọn tối ưu nói trên được xác định cùng với một giá trị của nhân tử Lagrange $\bar{\lambda} = \bar{\lambda}(b)$ và giá trị tối ưu của w là hàm số của b

$\bar{w} = f(x_1, x_2, \ldots, x_n) = \bar{w}(b)$

Theo quy tắc đạo hàm hợp ta có:

$$\frac{d\bar{w}}{db} = \frac{\delta \bar{f}}{\delta x_1}\frac{d\bar{x_1}}{db} + \frac{\delta \bar{f}}{\delta x_2}\frac{d\bar{x_2}}{db} + \cdots + \frac{\delta \bar{f}}{\delta x_n}\frac{d\bar{x_n}}{db}$$

Do $(\bar{\lambda}, \bar{x_1}, \bar{x_2}, \ldots \bar{x_n})$ là điểm dừng của hàm số Lagrange nên ta có:

$\frac{\delta \bar{f}}{\delta x_i} = \lambda \frac{\delta \bar{g}}{\delta x_i} \forall i \in 1, 2, \ldots n$

Vậy

$$\frac{d\bar{w}}{db} = \lambda \left\{ \frac{\delta \bar{g}}{\delta x_1}\frac{d\bar{x_1}}{db} + \frac{\delta \bar{g}}{\delta x_2}\frac{d\bar{x_2}}{db} + \cdots + \frac{\delta \bar{g}}{\delta x_n}\frac{d\bar{x_n}}{db} \right\} (1.4)$$

Mà ta lại có $g(x_1, x_2, \ldots, x_n) = 0$. Lấy đạo hàm hai vế của đồng nhất thức này theo b ta có:

$$\frac{\delta \bar{g}}{\delta x_1}\frac{d\bar{x_1}}{db} + \frac{\delta \bar{g}}{\delta x_2}\frac{d\bar{x_2}}{db} + \cdots + \frac{\delta \bar{g}}{\delta x_n}\frac{d\bar{x_n}}{db} = 1(1.5)$$

Từ (1.4) và (1.5) suy ra: $\bar{\lambda} = \frac{d\bar{w}}{db}(1.6)$

Hệ thức trên cho thấy nhân tử Lagrange $\bar{\lambda}$ chính là giá trị $\bar{w}$ cận biên của b. Điều này có nghãi là khi b tăng thêm 1 đơn vị thì giá trị tối ưu $\bar{w}$ của hàm mục tiêu thay đổi một lượng xấp xỉ bằng $\bar{\lambda}$.

Bài toán Chọn (x, y) để hàm số $w = f(x, y)$ đạt giá trị cực đại (cực tiểu) với điều kiện $g(x, y) \geq b$ [hoặc $g(x, y) \leq b$]

Phương pháp chung: Để giải bài toán này, trước hết ta thay điều kiện bằng phương trình $g(x, y) = b$. Bằng phương pháp nhân tử Lagrange ta tìm được $x = \bar{x}, y = \bar{y}$ và $\lambda = \bar{\lambda}$. Quyết định cuối cùng dựa theo nguyên tắc sau đây:

(a) Đối với bài toán cực đại hóa tham số $w = f(x, y)$ với điều kiện $g(x, y) \geq b$

- Nếu $\bar{\lambda} < 0$ thì điều kiện áp đặt thực sự là ràng buộc và phương án chọn tối ưu là $x = \bar{x}, y = \bar{y}$
- $\bar{\lambda} \geq 0$ thì điều kiện áp đặt không phải là ràng buộc thực sự. Trong trường hợp này ta bỏ điều kiện và quay sang giải bài toán cực trị không có điều kiện.

(b) Đối với bài toán cực đại hóa tham số $w = f(x, y)$ với điều kiện $g(x, y) \leq b$

- Nếu $\bar{\lambda} \geq 0$ thì điều kiện áp đặt thực sự là ràng buộc và phương án chọn tối ưu là $x = \bar{x}, y = \bar{y}$
- $\bar{\lambda} < 0$ thì điều kiện áp đặt không phải là ràng buộc thực sự. Trong trường hợp này ta bỏ điều kiện và quay sang giải bài toán cực trị không có điều kiện.

(c) Đối với bài toán cực tiểu hóa tham số $w = f(x, y)$ với điều kiện $g(x, y) \geq b$

- Nếu $\bar{\lambda} \geq 0$ thì điều kiện áp đặt thực sự là ràng buộc và phương án chọn tối ưu là $x = \bar{x}, y = \bar{y}$
- $\bar{\lambda} < 0$ thì điều kiện áp đặt không phải là ràng buộc thực sự. Trong trường hợp này ta bỏ điều kiện và quay sang giải bài toán cực trị không có điều kiện.

(d) Đối với bài toán cực tiểu hóa tham số $w = f(x, y)$ với điều kiện $g(x, y) \leq b$

- Nếu $\bar{\lambda} < 0$ thì điều kiện áp đặt thực sự là ràng buộc và phương án chọn tối ưu là $x = \bar{x}, y = \bar{y}$
- $\bar{\lambda} \geq 0$ thì điều kiện áp đặt không phải là ràng buộc thực sự. Trong trường hợp này ta bỏ điều kiện và quay sang giải bài toán cực trị không có điều kiện.

3. Support Vector Machine

Bài toán: Cho hai lớp và các điểm thuộc hai lớp đấy. Gọi các điểm đấy là các điểm dữ liệu. Cần tìm mặt phẳng d để khoảng cách giữa hai mặt phẳng song song, cách đều mặt phẳng d mà vùng không gian tạo bởi hai mặt phẳng đó vẫn không chứa bất kỳ điểm dữ liệu nào là lớn nhất.

Bài toán trên tương đương với bài toán:
Bài toán Cho hai lớp và các điểm thuộc hai lớp đấy. Gọi các điểm đấy là các điểm dữ liệu. Cần tìm hai mặt phẳng song song sao cho vùng không gian tạo bởi hai mặt phẳng đấy chia không gian còn lại thành hai phần khác nhau, mỗi phần chỉ chứa các điểm dữ liệu của một lớp mà khoảng cách giữa hai mặt phẳng đấy là lớn nhất.

Ta có: Bất kỳ hai mặt phẳng song song nào đều được biểu diễn bởi:

$$\vec{w}\vec{x} + c - 1 = 0 (d_1)$$

$$\vec{w}\vec{x} + c + 1 = 0 (d_2)$$

Khoảng cách giữa d_1 và d_2 là:

$$\Delta_{12} = \frac{2}{||\vec{w}||}$$

Giả sử tập các điểm dữ liệu là:

$$D = (\vec{x_1}, y_1), (\vec{x_2}, y_2), \ldots, (\vec{x_n}, y_n)$$

với n là số lượng dữ liệu.
Bài toán đã cho được viết lại thành:
Bài toán Cho tập $D = \{(\vec{x_1}, y_1), (\vec{x_2}, y_2), \ldots, (\vec{x_n}, y_n)\}$
Tìm giá trị của $\vec{w}, c$ để $\frac{||\vec{w}||^2}{2}$ nhỏ nhất với điều kiện:

$$\begin{cases} \vec{w}\vec{x_i} + c \geq 1, \text{nếu } y_i = 1 \\ \vec{w}\vec{x_i} + c \leq 1, \text{nếu } y_i = -1 \end{cases}$$

với $\forall (x_i, y_i) \in D$
Viết gọn điều kiện lại, ta có bài toán:
Bài toán Cho tập $D = \{(\vec{x_1}, y_1), (\vec{x_2}, y_2), \ldots, (\vec{x_n}, y_n)\}$
Tìm giá trị của $\vec{w}, c$ để $\frac{||\vec{w}||^2}{2}$ nhỏ nhất với điều kiện: $y_i(\vec{w}\vec{x_i} + c) \geq 1$
với $\forall (x_i, y_i) \in D$

Giải:
Xét hàm số Lagrange:

$$L(\vec{w}, c, \lambda) = \frac{||\vec{w}||^2}{2} + \sum_{i=1}^{n} \lambda_i [1 - y_i(\vec{w}\vec{x_i} + c)]$$

Viết λ thay thế cho n biến λ_i với $i \in 1 \ldots n$

Ta tiếp cận bài toán sau trước:
Cho tập $D = \{(\vec{x_1}, y_1), (\vec{x_2}, y_2), \ldots, (\vec{x_n}, y_n)\}$
Tìm giá trị của $\vec{w}, c$ để $\frac{||\vec{w}||^2}{2}$ nhỏ nhất với điều kiện: $y_i(\vec{w}\vec{x_i} + c) = 1$
với $\forall (x_i, y_i) \in D$
Ta tiếp cận bằng phương pháp Lagrange:
Giả sử các hàm số $f(\vec{w}, c) = \frac{||\vec{w}||}{2}$ và $g_i(\vec{w}, c) = y_i(\vec{w}\vec{x_i} + c) \forall i \in 1 \ldots n$ có đạo hàm riêng cấp hai liên tục. Khi đó nếu hàm số $f(\vec{w}, c)$, với điều kiện $g_i(\vec{w}, c) = 1$ đạt cực trị tại điểm $M_0(\vec{\bar{w}}, \bar{c})$ thì tồn tại bộ $(\vec{\bar{w}}, \bar{c}, \bar{\lambda})$ là nghiệm của hệ phương trình:

$$\begin{cases} L'_{\lambda_i} = 1 - y_i(\vec{w}\vec{x_i} + c) = 0 \forall i \in 1 \ldots n \\ L'_{\vec{w}} = \vec{w} - \sum_{i=1}^{n} \lambda_i y_i \vec{x_i} = 0 \\ L'_c = -\sum_{i=1}^{n} \lambda_i y_i = 0 \end{cases}$$

Quay lại bài toán ban đầu:
Giả sử $(\bar{\vec{w}}, \bar{c}, \bar{\lambda})$ là một nghiệm của hệ phương trình trên.
Xét các biến $\bar{\lambda}(\bar{\lambda}_1, \bar{\lambda}_2, \dots \bar{\lambda}_n)$
Giả sử P, Q là tập con của $\bar{\lambda}$ sao cho $\bar{\lambda}_i \geq 0 \forall \bar{\lambda}_i \in P$ và $\bar{\lambda}_i < 0 \forall \bar{\lambda}_i \in Q$

Theo ý nghĩa của nhân tử Lagrange trong phụ đề (∗∗) ta có $\forall \bar{\lambda}_i \in \bar{\lambda}$ ta có: $\bar{\lambda}_i = \frac{d\bar{f}}{db_i}$.
Vậy ta có $\forall \bar{\lambda}_i \in Q$ thì khi b_i tăng thì f sẽ giảm tại lân cận của b_i hay f không đạt giá trị nhỏ nhất tại biên $g_i(\vec{w}, c) = b_i$. Vậy ta sẽ bỏ qua điều kiện $g_i(\vec{w}, c) \geq b_i$ để tìm giá trị nhỏ nhất của f (điều kiện cần để tại điểm $M_0(\bar{\vec{w}}, \bar{c})$ f lớn nhât với điều kiện $g_i(\vec{w}, c) \geq b_i$ là tại $M_0(\bar{\vec{w}}, \bar{c})$ f lớn nhât không có điều kiện $g_i(\vec{w}, c) \geq b_i$)

Và ta cũng có $\forall \bar{\lambda}_i \in P$ thì khi b_i giảm thì f sẽ giảm tại lân cận của b_i hay f đạt giá trị nhỏ nhất với điều kiện $g_i(\vec{w}, c) \geq b_i$ tại điểm thỏa mãn điều kiện $g_i(\vec{w}, c) = b_i$

Theo đó để giải quyết bài tóan ban đầu, ta chuyển sang giải quyết bài toán sau:
Cho tập $D = \{(\vec{x_1}, y_1), (\vec{x_2}, y_2), \dots, (\vec{x_n}, y_n)\}$
và giả sử $\bar{\lambda}(\bar{\lambda}_1, \bar{\lambda}_2, \dots \bar{\lambda}_n)$ là điểm để f đạt cực tiểu với điều kiện $y_i(\vec{w}\vec{x_i} + c) \geq 1$ với $\forall (x_i, y_i) \in D$.
Giả sử P là tập con của $\bar{\lambda}$ sao cho $\bar{\lambda}_i \geq 0 \forall \bar{\lambda}_i \in P$

Giả sử tập $D_1 \in D$ sao cho $\bar{\lambda}_i$ tương ứng với $(\vec{x_i}, y_i)$ thỏa mãn thuộc tập P. Tìm giá trị của $\vec{w}, c$ để $\frac{||\vec{w}||^2}{2}$ nhỏ nhất với điều kiện: $y_i(\vec{w}\vec{x_i} + c) = 1$ với $\forall (x_i, y_i) \in D_1$

Xem tập D_1 như tập D ở trên, tương tự như trên ta có:
Xét hàm số Lagrange:

$$L(\vec{w}, c, \lambda) = \frac{||\vec{w}||^2}{2} + \sum_{\forall (x_i, y_i) \in D_1} \lambda_i [1 - y_i(\vec{w}\vec{x_i} + c)]$$

Giả sử các hàm số $f(\vec{w}, c) = \frac{||\vec{w}||}{2}$ và $g_i(\vec{w}, c) = y_i(\vec{w}\vec{x_i} + c) \forall (x_i, y_i) \in D_1$ có đạo hàm riêng cấp hai liên tục. Khi đó nếu hàm số $f(\vec{w}, c)$, với điều kiện $g_i(\vec{w}, c) = 1$ đạt cực trị tại điểm $M_0(\bar{\vec{w}}, \bar{c})$ thì tồn tại bộ $(\bar{\vec{w}}, \bar{c}, \bar{\lambda})$ là nghiệm của hệ phương trình:

$$\begin{cases} L'_{\lambda_i} = 1 - y_i(\vec{w}\vec{x_i} + c) = 0 \forall (x_i, y_i) \in D_1 \\ L'_{\vec{w}} = \vec{w} - \sum_{\forall (x_i, y_i) \in D_1} \lambda_i y_i \vec{x_i} = 0 \\ L'_c = -\sum_{\forall (x_i, y_i) \in D_1} \lambda_i y_i = 0 \end{cases}$$

Hay tương đương với:

$$\begin{cases} y_i(\vec{w}\vec{x_i} + c) = 1 \forall (x_i, y_i) \in D_1 \\ \vec{w} = \sum_{\forall (x_i, y_i) \in D_1} \lambda_i y_i \vec{x_i} \\ \sum_{\forall (x_i, y_i) \in D_1} \lambda_i y_i = 0 \end{cases}$$

Giả sử $(\bar{\vec{w}}, \bar{c}, \bar{\lambda})$ là một nghiệm của hệ phương trình trên.
Khi đó

$$L(\vec{w}, c, \lambda) = \frac{||\vec{w}||^2}{2}$$

$$\Leftrightarrow L(\vec{w}, c, \lambda) = \frac{(\sum_{\forall (x_i, y_i) \in D_1} \lambda_i y_i \vec{x_i})^2}{2} (2.1)$$

Ta sẽ tìm giá trị nhỏ nhất của (2.1) với điều kiện $\sum_{\forall (x_i, y_i) \in D_1} \lambda_i y_i = 0$ và $\lambda_i \geq 0 \forall \lambda_i \in P$