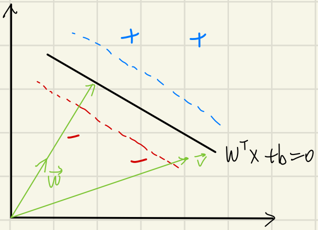


Step 1: Decision Rule

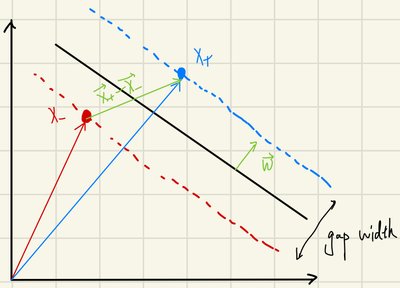
$$\boxed{\text{if } w^T v + b \geq 0 \text{ then } +}$$

$\Downarrow$ We want a more "confident" classification

want: $\begin{cases} w^T x_+ + b \geq 1 \\ w^T x_- + b \leq -1 \end{cases} \Rightarrow \boxed{y_i (w^T x_i + b) \geq 1}$

Step 2: "confident" decision rule

$y_i (w^T x_i + b) = 1$ if x is on the dash line



$$\text{width} = (x_+ - x_-)^T w \,/\, \|w\|_2$$

$$= [1 - b - (-1 - b)] \,/\, \|w\|_2$$

$$= 2 / \|w\|_2$$

$$\max\ 2/\|w\|_2 \Rightarrow \min\ \|w\|_2 \Rightarrow \min\ \tfrac{1}{2}\|w\|_2^2$$

$$\boxed{\begin{aligned} &\min.\ \ 1/2\, \|w\|_2^2 \\ &\text{s.t.}\ \ y_i (w^T x + b) \geq 1 \end{aligned}}$$

Step 3: Formulate the problem as an optimization problem

(which is a standard-form convex optimization problem and can be solved using Interior-point method)

Step 4

Lagrangian is our tool to solve the problem

$$\begin{aligned} &\min.\ \ f_0(x) \\ &\text{s.t.}\ \ f_i(x) \leq 0 \\ &\qquad\ h_i(x) = 0 \end{aligned}$$

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_i \lambda_i f_i(x) + \sum_i \nu_i h_i(x)$$

$$g(x, \nu) = \inf_{x \in D} \mathcal{L}(x, \lambda, \nu)$$

dual: $\max.\ g(x, \nu)$

$\quad$ s.t. $\lambda_i \geq 0$

Assume strong duality holds, x^* primal optimal, λ^*, ν^* is dual optimal

$$f_0(x^*) = g(\lambda^*, \nu^*) = \inf_{x \in D} \left[f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_i \nu_i^* h_i(x) \right]$$

$$\leq f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_i \nu_i^* h_i(x)$$

$$= \leftarrow \leq f_0(x^*)$$

KKT $\begin{cases} \text{primal feasible: } f_i(x) \leq 0 \quad h_i(x) = 0 \\ \text{dual feasible: } \lambda_i \geq 0 \\ \text{complementary slackness: } \lambda_i f_i(x) = 0 \\ \text{gradient vanishes: } \nabla_x \mathcal{L}(x^*, \lambda^*, \nu^*) = 0 \end{cases}$

for a convex problem

KKT $\Leftrightarrow$ primal-dual-optimal

• Lagrangian

$$\min.\ \frac{1}{2}\|W\|_2^2$$
$$\text{s.t. } 1-y_i(W^Tx_i+b)\le 0 \quad (\text{dual variable } \alpha)$$

$$\mathcal{L}(w,b,\alpha)=\frac{1}{2}\|W\|_2^2-\sum_i\alpha_i\left[y_i(W^Tx_i+b)-1\right]$$
$$=\frac{1}{2}\|W\|_2^2-\sum_i\alpha_i y_i W^Tx_i - b\sum_i\alpha_i y_i+\sum_i\alpha_i$$
$$=\frac{1}{2}\|W\|_2^2-W^T\left(\sum_i\alpha_i y_i x_i\right)-b\sum_i\alpha_i y_i+\sum_i\alpha_i$$

$\mathcal{L}$ is quadratic in w, affine in b

$$\nabla_w\mathcal{L}=w-\sum_i\alpha_i y_i x_i := 0$$
$$W=\sum_i\alpha_i y_i x_i$$

$$g(\alpha)=\inf_{w,b}\mathcal{L}(w,b,\alpha)=\begin{cases}-\infty & \sum_i\alpha_i y_i\neq 0\\ -\frac{1}{2}\left(\sum_i\alpha_i y_i x_i\right)^T\left(\sum_i\alpha_i y_i x_i\right)+\sum_i\alpha_i & \end{cases}$$

$\Downarrow$

dual problem

$$\max.\ -\frac{1}{2}\sum_i\sum_j\alpha_i\alpha_j y_i y_j x_i^Tx_j+\sum_i\alpha_i$$
$$\text{s.t. } \alpha_i\ge 0$$
$$\sum_i\alpha_i y_i=0$$

Lagrangian (relaxed)

$$\min.\ \frac{1}{2}\|W\|_2^2+C\cdot\sum_i\varepsilon_i$$
$$\text{s.t. } 1-\varepsilon_i-y_i(W^Tx_i+b)\le 0 \quad (\text{dual variable } \alpha)$$
$$-\varepsilon_i\le 0 \quad (\text{dual variable } \beta)$$

$$\mathcal{L}(w,b,\varepsilon,\alpha,\beta)$$
$$=\frac{1}{2}\|W\|_2^2+C\sum_i\varepsilon_i-\sum_i\alpha_i\left[y_i(W^Tx_i+b)+\varepsilon_i-1\right]-\sum_i\beta_i\varepsilon_i$$
$$=\frac{1}{2}\|W\|_2^2-W^T\sum_i\alpha_i y_i x_i-b\sum_i\alpha_i y_i+\sum_i\alpha_i$$
$$+\sum_i(C-\alpha_i-\beta_i)\varepsilon_i$$

$\mathcal{L}$ is quadratic in w, affine in b, ε_i

$$\nabla_w\mathcal{L}=W-\sum_i\alpha_i y_i x_i := 0 \qquad W=\sum_i\alpha_i y_i x_i$$

$$g(\alpha\ \beta)=\inf_{w,b,\varepsilon}\mathcal{L}(w,b,\varepsilon,\alpha,\beta)$$
$$=\begin{cases}-\infty & \sum_i\alpha_i y_i\neq 0\\ -\infty & C-\alpha_i-\beta_i\neq 0\\ -\frac{1}{2}\left(\sum_i\alpha_i y_i x_i\right)^T\left(\sum_i\alpha_i y_i x_i\right)+\sum_i\alpha_i & \end{cases}$$

$\Downarrow$

dual problem

$$\max.\ -\frac{1}{2}\sum_i\sum_j\alpha_i\alpha_j y_i y_j x_i^Tx_j+\sum_i\alpha_i$$
$$\text{s.t. } \alpha_i\ge 0 \quad \beta_i\ge 0$$
$$\sum_i\alpha_i y_i=0 \quad C-\alpha_i-\beta_i=0$$

$\Downarrow$

$$\max\ -\frac{1}{2}\sum_i\sum_j\alpha_i\alpha_j y_i y_j x_i^Tx_j+\sum_i\alpha_i$$
$$\text{s.t. } \sum_i\alpha_i y_i=0$$
$$0\le\alpha_i\le C$$

• KKT Condition

primal problem

$$\min. \ \frac{1}{2}\|W\|_2^2$$

$$\text{s.t.} \quad y_i(W^T x_i + b) - 1 \geq 0 \qquad \alpha_i$$

dual problem

$$\max \ -\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i$$

$$\text{s.t.} \quad \sum_i \alpha_i y_i = 0$$

$$\alpha_i \geq 0$$

KKT Condition:

primal feasible: $y_i(W^T x_i - b) - 1 \geq 0$

dual feasible: $\sum_i \alpha_i y_i = 0$

$\alpha_i \geq 0$

complemetary slackness:

$$\alpha_i\left[y_i(w^T x_i + b) - 1\right] = 0$$

gradiend of $\mathcal{L}$

$$W = \sum_i \alpha_i y_i x_i$$

primal problem

$$\min. \ \frac{1}{2}\|W\|_2^2 + C\sum_i \xi_i$$

$$\text{s.t.} \quad y_i(W^T x_i + b) - 1 + \xi_i \geq 0 \qquad \alpha_i$$

$$\xi_i \geq 0 \qquad \beta_i$$

dual problem

$$\max. \ -\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i$$

$$\text{s.t.} \quad \sum_i \alpha_i y_i = 0$$

$$\alpha_i \geq 0 \quad \beta_i \geq 0 \quad C - \alpha_i - \beta_i = 0$$

KKT Condition:

primal feasible: $y_i(W^T x_i + b) - 1 + \xi_i \geq 0$

$\xi_i \geq 0$

dual feasible: $\sum_i \alpha_i y_i = 0$

$$0 \leq \alpha_i \leq C \rightarrow \begin{cases} \alpha_i \geq 0 \\ \beta_i \geq 0 \\ C - \alpha_i - \beta_i = 0 \end{cases}$$

complementary slackness:

if $\alpha_i = 0 \quad \beta_i = C > 0 \quad \xi_i = 0$

$y_i(W^T x_i + b) \geq 1 - \xi_i = 1$

if $\alpha_i = C \quad \beta_i = 0 \quad \xi_i > 0$

$y_i(W^T x_i + b) = 1 - \xi_i \leq 1$

if $0 < \alpha_i < C \quad \beta > 0 \quad \xi_i = 0$

$y_i(W^T x_i + b) = 1 - \xi_i = 1$

gradient of $\mathcal{L}$:

$$W = \sum_i \alpha_i y_i x_i$$

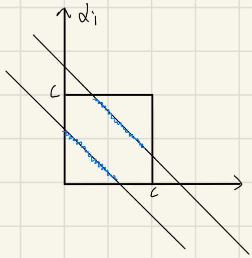
· SMO

select α_i and α_j dual feasible : $\sum_i \alpha_i y_i = 0$

$$\alpha_i y_i + \alpha_j y_j = -\sum_{\substack{k\neq i \\ k\neq j}} \alpha_k y_k$$

1. when $y_i = y_j$

$$\alpha_i^{new} + \alpha_j^{new} = \alpha_i^{old} + \alpha_j^{old} \geq 0$$

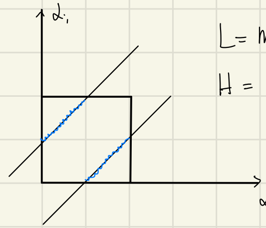


$L = \max\{0, \alpha_i^{old} + \alpha_j^{old} - C\}$

$H = \min\{C, \alpha_i^{old} + \alpha_j^{old}\}$

2. when $y_i \neq y_j$

$$\alpha_i^{new} - \alpha_j^{new} = \alpha_i^{old} - \alpha_j^{old}$$



$L = \max\{0, \alpha_i^{old} - \alpha_j^{old}\}$

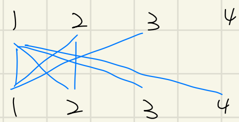
$H = \min\{C, \alpha_i^{old} - \alpha_j^{old} + C\}$

dual problem :

$$\max. \ -\frac{1}{2}\sum_i\sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i$$

$$\text{s.t.} \ \sum \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C$$



$$W(\alpha_1, \alpha_2) = -\frac{1}{2}\alpha_1^2 x_1^T x_1 - \frac{1}{2}\alpha_2^2 x_2^T x_2 - \alpha_1 \alpha_2 y_1 y_2 x_1^T x_2$$

$$-(\alpha_1 y_1 x_1)^T \sum_{i=3}(\alpha_i y_i x_i) - (\alpha_2 y_2 x_2)^T \sum_{i=3}(\alpha_i y_i x_i) + \alpha_1 + \alpha_2$$

$$\underbrace{-\frac{1}{2}\sum_{i=3}\sum_{j=3} \alpha_i \alpha_j y_i y_j (x_i^T x_j) + \sum_{i=3} \alpha_3}_{\text{constant}}$$

$$\alpha_2 = \frac{D - \alpha_1 y_1}{y_2} \qquad \alpha_1 y_1 + \alpha_2 y_2 = D$$

$$W(\alpha_1) = -\frac{1}{2}\alpha_1^2 x_1^T x_1 - \frac{1}{2}(D - \alpha_1 y_1)^2 x_2^T x_2 - \alpha_1 \frac{D - \alpha_1 y_1}{y_2} y_1 y_2 x_1^T x_2$$

$$-\alpha_1 y_1 \sum_{i=3} \alpha_i y_i x_i^T x_1 - (D - \alpha_1 y_1) \sum_{i=3} \alpha_i y_i x_i^T x_2 + \alpha_1 + \frac{D - \alpha_1 y_1}{y_2} + \text{constant}$$

$$\frac{dW}{d\alpha_1} = -\alpha_1 x_1^T x_1 + (D - \alpha_1 y_1) y_1 x_2^T x_2 - D y_1 x_1^T x_2 + 2\alpha_1 x_1^T x_2$$

$$- y_1 \sum_{i=3} \alpha_i y_i x_i^T x_1 + y_1 \sum_{i=3} \alpha_i y_i x_i^T x_2 + 1 - \frac{y_1}{y_2}$$

$$= -\alpha_1 x_1^T x_1 - \alpha_1 x_2^T x_2 + 2\alpha_1 x_1^T x_2$$

$$+ D y_1 x_2^T x_2 - D y_1 x_1^T x_2 - y_1 \sum_{i=3} \alpha_i y_i x_i^T x_1 + y_1 \sum_{i=3} \alpha_i y_i x_i^T x_2 + 1 - \frac{y_1}{y_2}$$

$$= -\alpha_1 x_1^T x_1 - \alpha_1 x_2^T x_2 + 2\alpha_1 x_1^T x_2$$
$$+ (\alpha_1^{old} y_1 + \alpha_2^{old} y_2)\, y_1 (x_2^T x_2 - x_1^T x_2) + y_1 \left[\sum_{i=3} \alpha_i y_i x_i^T (x_2 - x_1) \right] + 1 - \frac{y_1}{y_2}$$

$$:= 0$$

$$\alpha_1 = \frac{(\alpha_1^{old} y_1 + \alpha_2^{old} y_2)\, y_1 (x_2^T x_2 - x_1^T x_2) + y_1 \left[\sum_{i\geq 3} \alpha_i y_i x_i^T (x_2 - x_1) \right] + 1 - \frac{y_1}{y_2}}{x_1^T x_1 + x_2^T x_2 - 2x_1^T x_2}$$

$$= y_1 \cdot \frac{(\alpha_1^{old} y_1 + \alpha_2^{old} y_2)(x_2^T x_2 - x_1^T x_2) + \sum_{i\geq 3} \alpha_i y_i x_i^T x_2 - \sum_{i\geq 3} \alpha_i y_i x_i^T x_1 + y_1 - y_2}{x_1^T x_1 + x_2^T x_2 - 2x_1^T x_2}$$

$$= y_1 \frac{y_1 \alpha_1^{old} x_2^T x_2 - y_1 \alpha_1^{old} x_1^T x_2 + \alpha_2^{old} y_2 x_2^T x_2 - \alpha_2^{old} y_2 x_1^T x_2 + \sum_{i\geq 3} \alpha_i y_i x_i^T x_2 - \sum_{i\geq 3} \alpha_i y_i x_i^T x_1 + y_1 - y_2}{x_1^T x_1 + x_2^T x_2 - 2x_1^T x_2}$$

$$= y_1 \frac{y_1 \alpha_1^{old} x_2^T x_2 - y_1 \alpha_1^{old} x_1^T x_2 - \alpha_1^{old} y_1 x_1^T x_2 + w^{old} x_2 + \alpha_1^{old} y_1 x_1^T x_1 - w^{old} x_1 + y_1 - y_2}{x_1^T x_1 + x_2^T x_2 - 2x_1^T x_2}$$

$$= y_1 \frac{(w^{old} x_2 - y_2) - (w^{old} x_1 - y_1) + \alpha_1^{old} y_1 x_1^T x_1 + \alpha_1^{old} y_1 x_2^T x_2 - 2\alpha_1^{old} y_1 x_1^T x_2}{x_1^T x_1 + x_2^T x_2 - 2x_1^T x_2}$$

$$= y_1 \left[\frac{(w^{old} x_2 - y_2) - (w^{old} x_1 - y_1)}{x_1^T x_1 + x_2^T x_2 - 2x_1^T x_2} + \alpha_1^{old} y_1 \right]$$

$$= \alpha_1^{old} + y_1 \frac{(w^{old} x_2 - y_2) - (w^{old} x_1 - y_1)}{x_1^T x_1 + x_2^T x_2 - 2x_1^T x_2}$$