Hard margin support vector machine

Yan PEI/裴岩 /ペイ　イエン

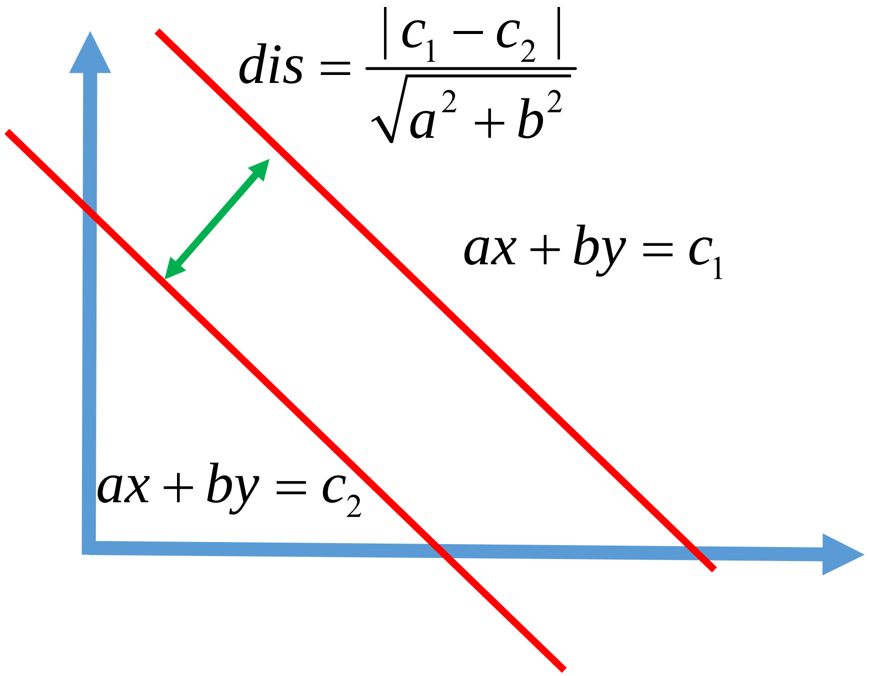
peiyan@u-aizu.ac.jp

Review some concepts and theorems

- 1 Langrange
- 2 dual form and duality theorem
- 3 KKT
- 4 kernel, feature space, kernel function

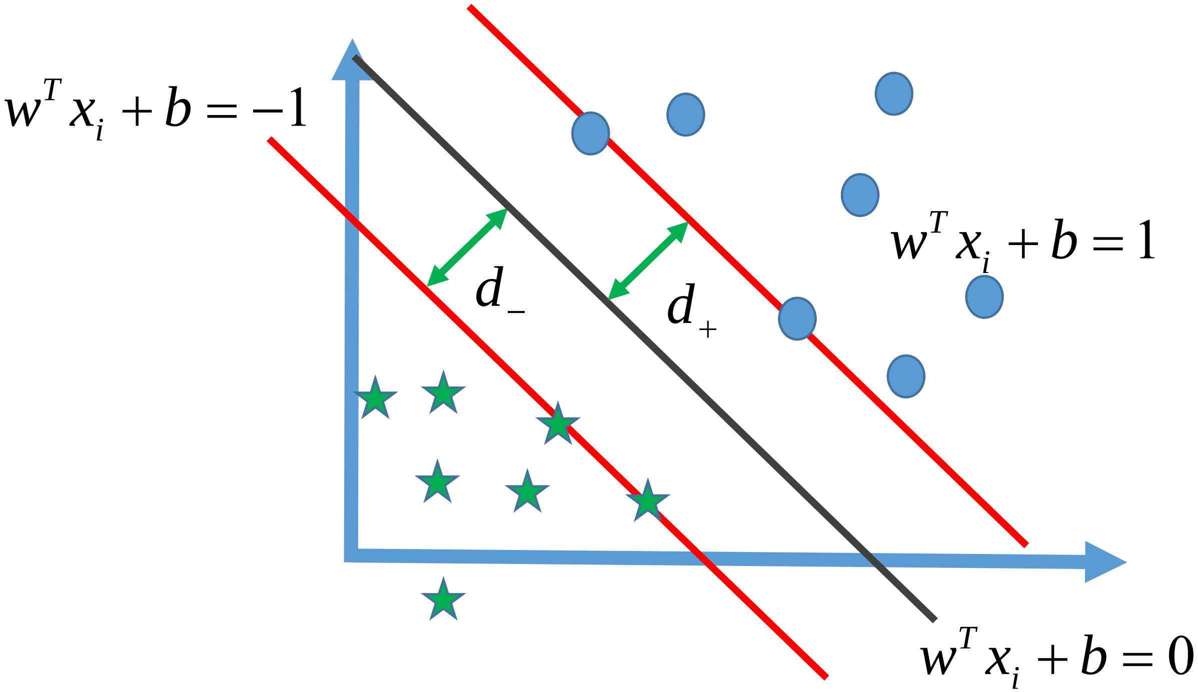
Review some concepts and theorems

- Distance between two parallel lines



Hard margin SVM

- Basic concepts and objective function

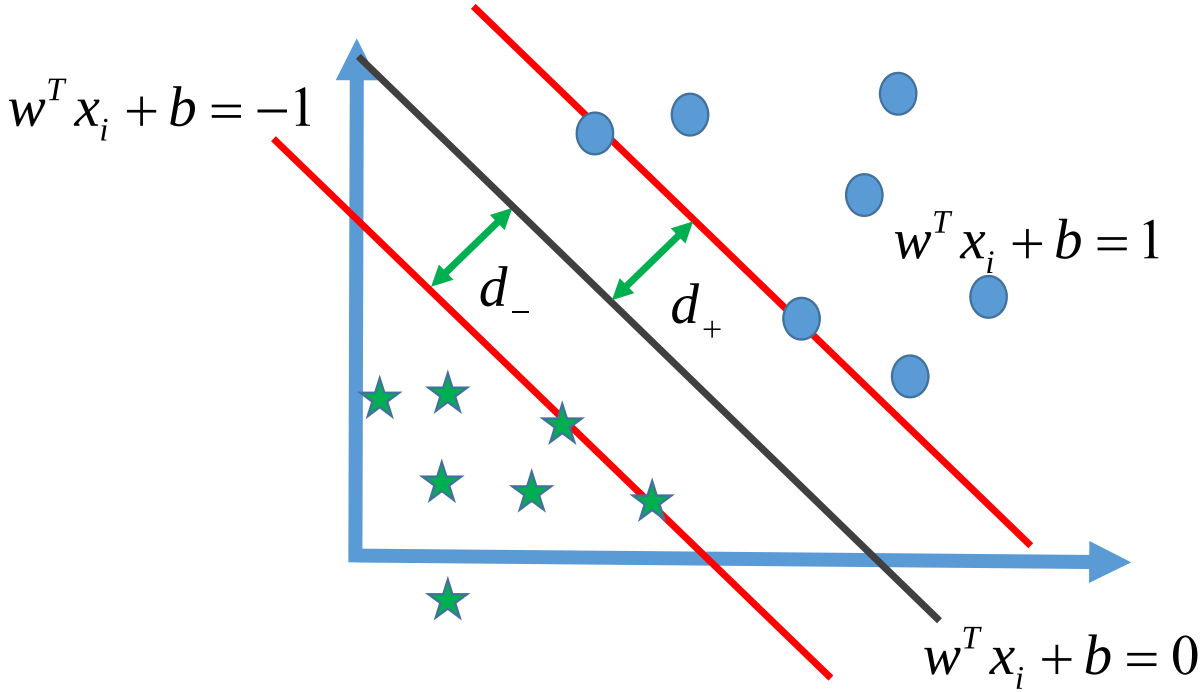


$$margin = d_+ + d_-$$

$$objective : \max(margin)$$

Hard margin SVM

- Dual form of objective function



Primal problem

$$w^T x_i + b \geq 1, if : y = 1$$

$$w^T x_i + b \leq -1, if : y = -1$$

$$margin : d_+ + d_-$$

$$y_i(w^T x_i + b) \geq 1$$

$$d_+ + d_- = \frac{1}{|w|} + \frac{1}{|w|} = \frac{2}{\sqrt{w^T w}}$$

$$subject\_to$$

$$y_i(w^T x_i + b) \geq 1$$

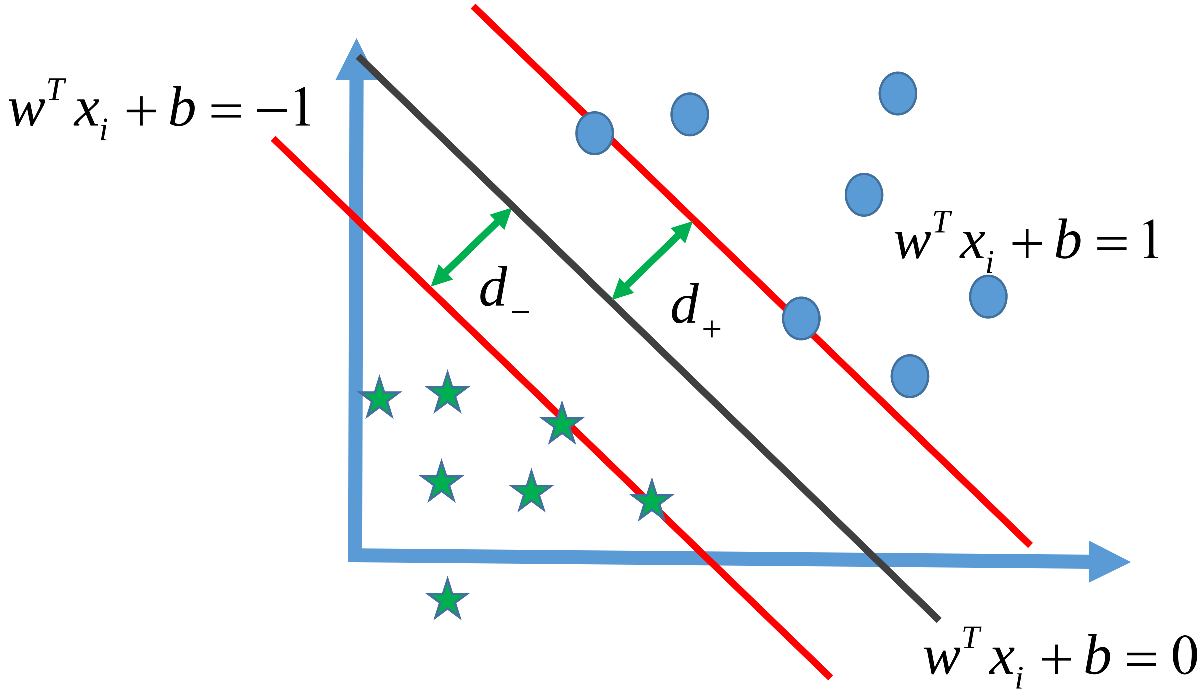
$$\min :$$

$$\varphi(w) = \frac{1}{2} w^T w$$

derivation

Hard margin SVM

- Dual form of objective function



$$subject\_to$$
$$y_i(w^T x_i + b) \geq 1$$
$$\min:$$
$$\varphi(w) = \frac{1}{2} w^T w$$

Dual problem

$$L(w,b,\alpha)$$
$$= \frac{1}{2} w^T w - \sum_{i=1}^{l} \alpha_i [y_i(w^T x_i + b) - 1]$$
$$\frac{\partial L}{\partial w} = \ldots = 0 \rightarrow w(\alpha) = \sum_{i=1}^{l} \alpha_i y_i x_i$$
$$\frac{\partial L}{\partial b} = \ldots = 0 \rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0$$

derivation

Hard margin SVM

- Now, we project samples into feature space and solve it by
  - Lagrange multipliers,
  - duality theorem
  - KKT

$$\min : L(w,b,\alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{l} \alpha_i [y_i (w^T \varphi(x_i) + b) - 1]$$

$$s.t.$$

$$\alpha_i \geq 0$$

$$\frac{\partial L}{\partial w} = \ldots = 0$$

$$\frac{\partial L}{\partial b} = \ldots = 0$$

derivation

Hard margin SVM

- Dual form/dual problem

$$W(\alpha) = L(w(\alpha), b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{l} \alpha_i [y_i (w^T \varphi(x_i) + b) - 1]$$

$$\frac{\partial L}{\partial w} = \ldots = 0 \rightarrow w(\alpha) = \sum_{i=1}^{l} \alpha_i y_i \varphi(x_i)$$

$$\frac{\partial L}{\partial b} = \ldots = 0 \rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0$$

$\max : \alpha\_unknown$

$$W(\alpha) = L(w(\alpha), b, \alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \kappa(x_i x_j)$$

$s.t.$

$$\sum_{i=1}^{l} \alpha_i y_i = 0$$

$$\alpha_i \geq 0$$

Karush-Kuhn-Tucker Theorem (KKT)

$$Let: f, g, h \in C$$

$$Let: x^{*} \_be\_a\_regular\_point$$

$$and\_a\_local\_minimizer\_for$$

$$problem\_f\_subject\_to$$

$$h(x) = 0, g(x) \le 0$$

$$Then, there\_exist\_\lambda^{*} \in R^{m} and\, \alpha^{*} \in R^{p}$$

$$(1)\alpha^{*} \ge 0$$

$$(2)\nabla f(x^{*}) + \lambda^{*T}\nabla h(x^{*}) + \alpha^{*T}\nabla g(x^{*}) = 0^{T}$$

$$(3)\alpha^{*T} g(x^{*}) = 0$$

$$(3.1)\alpha^{*} > 0 \to g(x^{*}) = 0$$

$$(3.2)g(x^{*}) < 0 \to \alpha^{*} = 0$$

derivation

Hard margin SVM

- KKT: it indicates that the Lagrange parameters can be non-zero only if the corresponding inequality is an equality at the solution.

Primal problem

$$\min:$$

$$\varphi(w)=\frac{1}{2}w^T w$$

$$s.t.$$

$$y_i(w^T x_i+b)\geq 1$$

Dual problem

$$\max: \alpha\_unknown$$

$$W(\alpha)=L(w(\alpha),b,\alpha)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j \kappa(x_i x_j)$$

$$s.t.$$

$$\sum_{i=1}^{l}\alpha_i y_i=0$$

$$\alpha_i\geq 0$$

Hard margin SVM --support vectors

- $From\_KKT$

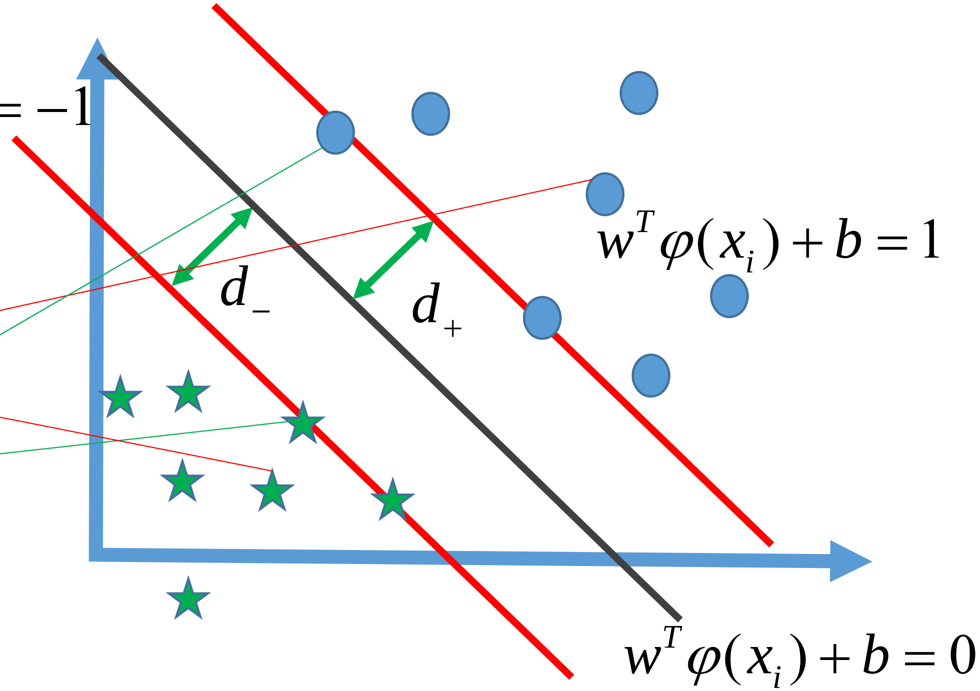
  $if:(\alpha^*, w^*, b^*) = optimal\_solution$

  $then: \alpha_t^*(y_t(w^{*T}\varphi(x_t)+b^*)-1)=0$

Primal problem

$$\min: \quad \varphi(w)=\frac{1}{2}w^T w$$

$$s.t. \quad y_i(w^T x_i + b) \geq 1$$



$$\alpha_t^*(y_t(w^{*T}\varphi(x_t)+b^*)-1)=0$$

$$if: y_t(w^{*T}\varphi(x_t)+b^*)>1 \rightarrow \alpha_t^*=0$$

$$if: \alpha_t^* \neq 0 \rightarrow y_t(w^{*T}\varphi(x_t)+b^*)=1$$

$$w(\alpha)=\sum_{i=1}^{l}\alpha_i y_i x_i$$

derivation

Hard margin SVM

- Find b by KKT

$$from\_KKT: \alpha_t \geq 0$$

$$\because \alpha_t (y_t(w^T\varphi(x_t)+b)-1)=0$$

$$\therefore w^T\varphi(x_t)+b=\pm 1{=}y_t, if\ :\varphi(x_i)\_is\_on\_margin$$

$$\therefore b = y_t - w^T\varphi(x_t) = y_t - \sum_{j=1}^{l}\alpha_j y_j \kappa(x_t x_j)$$

Note: Some of the case, b can be computed by average value of corresponding alpha, when alpha more than zero

derivation

Hard margin SVM

- summary

$$w^T \varphi(x_{test}) + b = \sum_{j=1}^{l} \alpha_j y_j \kappa(x_{test} x_j) + b$$

$$y_{test} = sig(\sum_{j=1}^{l} \alpha_j y_j \kappa(x_{test} x_j) + b)$$