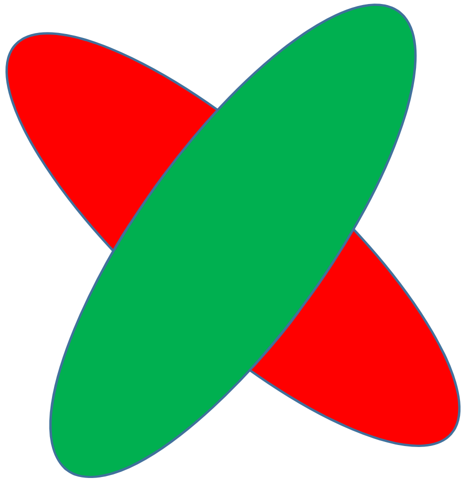
Soft margin support vector machine

Yan PEI/裴岩 /ペイ　イエン

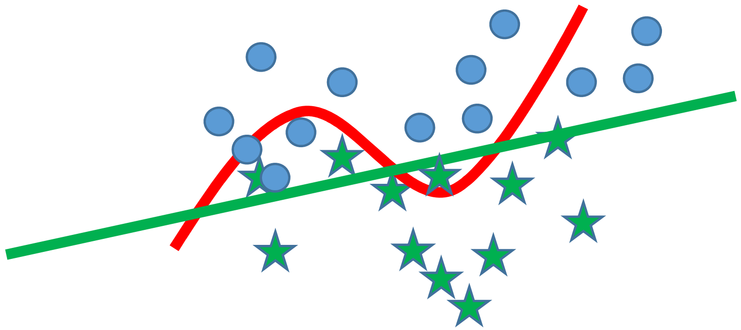
peiyan@u-aizu.ac.jp

Problems of SVM

- A linear hyper plane not exist

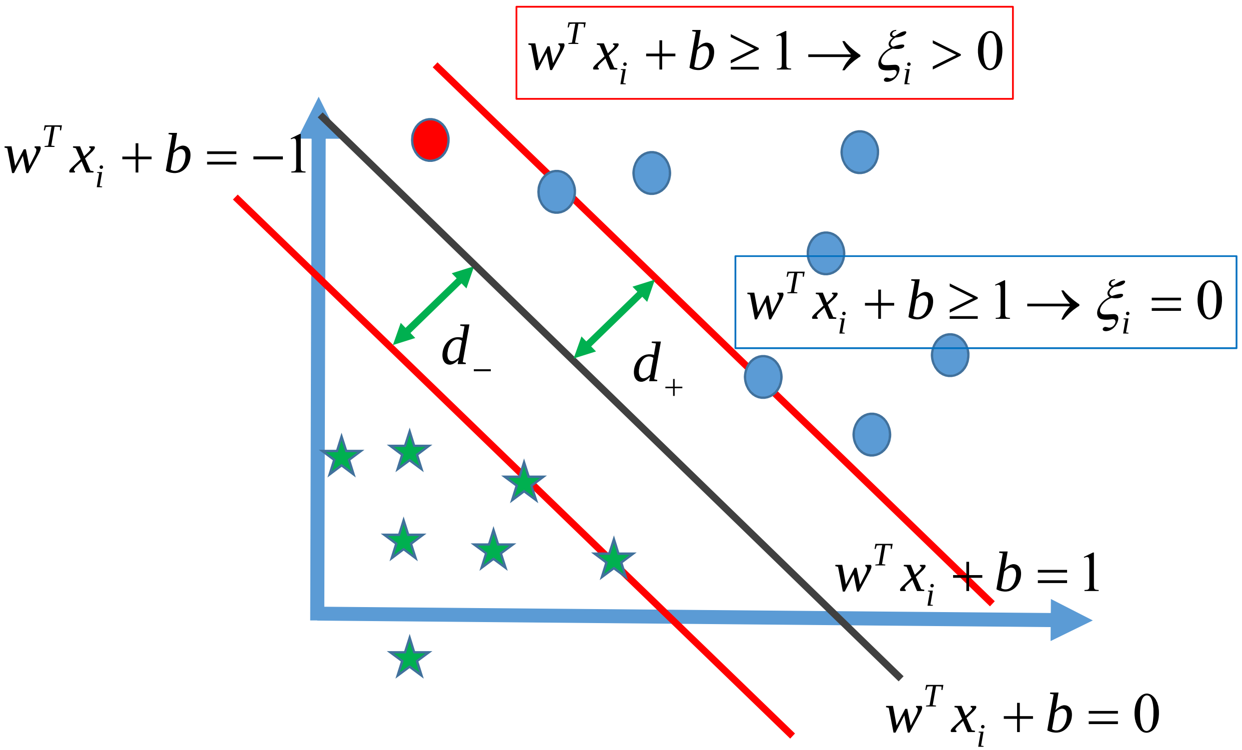


- Over fitting



soft margin SVM

- Dual form of objective function



$$subject\_to$$

$$y_i(w^T x_i + b) \ge 1\text{-}\xi_i$$

$$\min:$$

$$\varphi(w) = \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$

$$C > 0, \xi_i \ge o$$

Slack variables

derivation

Lagrange multipliers

- To solve the target function

$$subject\_to$$
$$y_i(w^T x_i + b) \geq 1\text{-}\xi_i$$
$$\xi_i \geq o$$

Slack variables

$$\min:$$
$$\varphi(w) = \frac{1}{2} w^T w + C\sum_{i=1}^{l} \xi_i$$
$$C > 0$$

$$L(w,b,\xi,\alpha,\beta)$$
$$= \frac{1}{2} w^T w + C\sum_{i=1}^{l} \xi_i - \sum_{i=1}^{l} \alpha_i [y_i (w^T x_i + b) - 1 + \xi_i] - \sum_{i=1}^{l} \beta_i \xi_i$$
$$\frac{\partial L}{\partial w} = \ldots = 0 \rightarrow w(\alpha) = \sum_{i=1}^{l} \alpha_i y_i x_i$$
$$\frac{\partial L}{\partial b} = \ldots = 0 \rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0$$
$$\frac{\partial L}{\partial \xi_i} = \ldots = 0 \rightarrow C - \alpha_i - \beta_i = 0 \rightarrow \begin{cases} \alpha_i = C - \beta_i \\ \beta_i = C - \alpha_i \end{cases}$$

derivation

Lagrange multipliers

- Now, we project x into a feature space

$$
\begin{aligned}
&subject\_to \\
&y_i(w^T\varphi(x_i)+b) \geq 1\text{-}\xi_i \\
&\xi_i \geq o \\
&\min: \\
&\varphi(w) = \frac{1}{2}w^T w + C\sum_{i=1}^{l}\xi_i \\
&C > 0
\end{aligned}
$$

Slack variables

$$
\begin{aligned}
&L(w,b,\xi,\alpha,\beta) \\
&= \frac{1}{2}w^T w + C\sum_{i=1}^{l}\xi_i - \sum_{i=1}^{l}\alpha_i[y_i(w^T\phi(x_i)+b)-1+\xi_i] - \sum_{i=1}^{l}\beta_i\xi_i \\
&\frac{\partial L}{\partial w} = \ldots = 0 \rightarrow w(\alpha) = \sum_{i=1}^{l}\alpha_i y_i \phi(x_i) = X^T Y\alpha \\
&\frac{\partial L}{\partial b} = \ldots = 0 \rightarrow \sum_{i=1}^{l}\alpha_i y_i = 0 = y^T\alpha \\
&\frac{\partial L}{\partial \xi_i} = \ldots = 0 \rightarrow C - \alpha_i - \beta_i = 0 \rightarrow \begin{cases} \alpha_i = C - \beta_i \\ \beta_i = C - \alpha_i \end{cases}
\end{aligned}
$$

derivation

Lagrange multipliers

- Now, we project x into a feature space

$$L(w,b,\xi,\alpha,\beta) = \frac{1}{2}w^T w + C\sum_{i=1}^{l}\xi_i - \sum_{i=1}^{l}\alpha_i[y_i(w^T\phi(x_i) + b) - 1 + \xi_i] - \sum_{i=1}^{l}\beta_i\xi_i$$

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

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

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

$$w^T w = [X^T Y\alpha]^T X^T Y\alpha$$

$$= \alpha^T Y^T X X^T Y\alpha = \alpha^T Y^T K Y\alpha = \sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j \kappa(x_i, x_j)$$

$$w^T\varphi(x_i) = [X^T Y\alpha]^T\varphi(x_i)$$

$$= \alpha^T Y^T X\varphi(x_i) = \sum_{j=1}^{l}\alpha_j y_j \kappa(x_i, x_j)$$

derivation

Dual problem

- Now, what is the lagrange multiplier? Dual problem

$$L(w,b,\xi,\alpha,\beta)=\frac{1}{2}w^T w+C\sum_{i=1}^{l}\xi_i-\sum_{i=1}^{l}\alpha_i[y_i(w^T\phi(x_i)+b)-1+\xi_i]-\sum_{i=1}^{l}\beta_i\xi_i$$

$$=\alpha^T 1_{l*1}-\frac{1}{2}\alpha^T YKY\alpha$$

$$w^T w=[X^T Y\alpha]^T X^T Y\alpha$$

$$=\alpha^T Y^T XX^T Y\alpha=\alpha^T Y^T KY\alpha=\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j\kappa(x_i,x_j)$$

$$w^T\varphi(x_i)=[X^T Y\alpha]^T\varphi(x_i)$$

$$=\alpha^T YX\varphi(x_i)=\sum_{j=1}^{l}\alpha_j y_j\kappa(x_i,x_j)$$

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

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

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

derivation

Soft margin SVM

- To solve this problem and find alpha, then we can obtain w and w*phi(x)

$$\max: W(\alpha) = \alpha^T 1_{l*1} - \frac{1}{2}\alpha^T YKY\alpha$$
$$subject\_to$$
$$(1)\alpha^T y = 0$$
$$(2) 0 \le \alpha_i \le C$$

$$w(\alpha) = \sum_{i=1}^{l} \alpha_i y_i \phi(x_i) = X^T Y\alpha$$
$$w(\alpha)\phi(z) = \sum_{i=1}^{l} \alpha_i y_i \phi(x_i)\phi(z) = \sum_{i=1}^{l} \alpha_i y_i \kappa(x_i, z)$$

KKT Explanation

KKT states that the Langrange parameters can be non-zero inly if the corresponding inequality is ***an equality*** at the solution

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

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

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

$$problem\_f\_subject\_to$$

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

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

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

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

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

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

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

$$L(w, b, \xi, \alpha, \beta) =$$

$$\frac{1}{2} w^{T} w + C \sum_{i=1}^{l} \xi_i - \sum_{i=1}^{l} \alpha_i [y_i (w^{T} \phi(x_i) + b) - 1 + \xi_i] - \sum_{i=1}^{l} \beta_i \xi_i$$

$$if: \alpha, w, b, \xi \in best\_solution$$

$$then$$

$$\alpha_i [y_i (w^{T} \phi(x_i) + b) - 1 + \xi_i] = 0$$

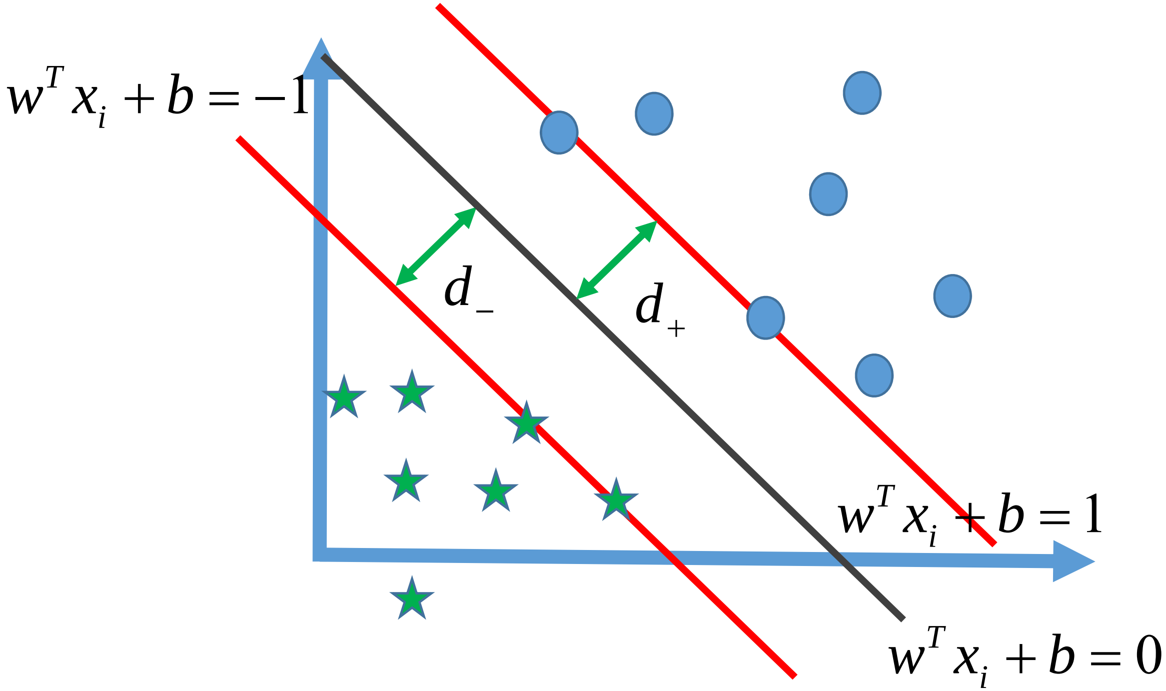
$$\beta_i \xi_i = 0$$

Let us discuss the parameter and its point distribution!!!

KKT Explanation

$$\alpha_i [y_i (w^T \phi(x_i) + b) - 1 + \xi_i] = 0$$
$$\beta_i \xi_i = 0$$
$$0 \le \alpha_i \le C$$

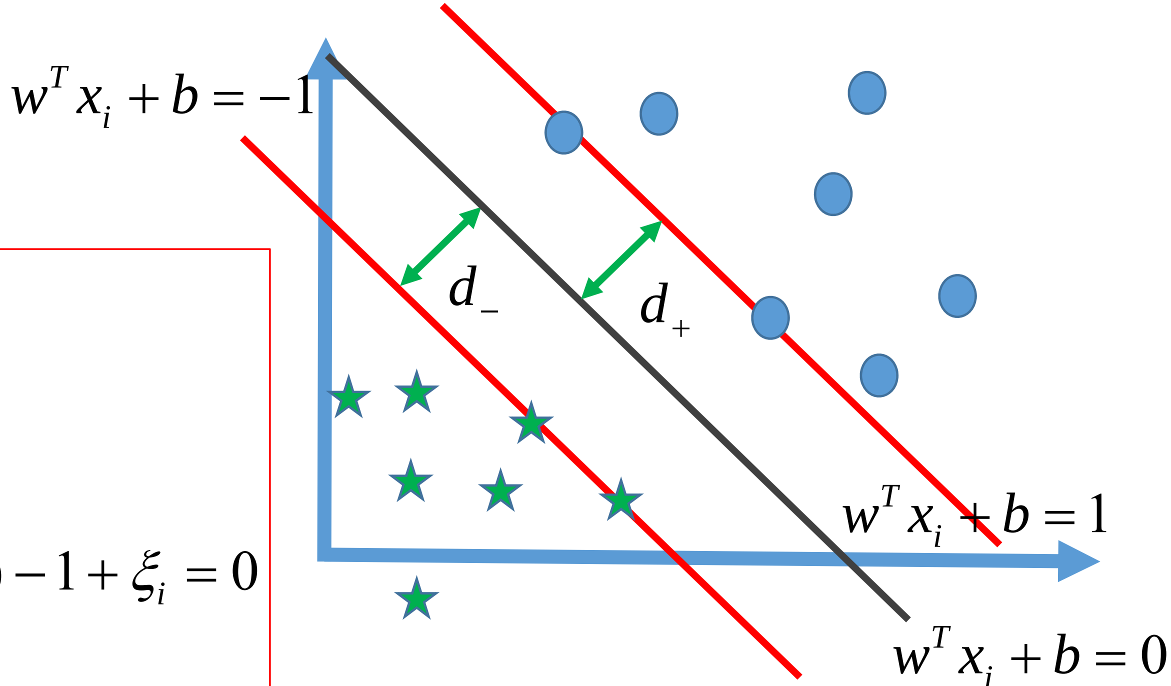
$$0 \le \alpha_i \le C$$
$$if\ : \alpha_i = 0$$
$$then:$$
$$\xrightarrow{\beta_i = C - \alpha_i} \beta_i = C$$
$$\xrightarrow{C>0 \& \beta_i \xi_i = 0} \xi_i = 0$$
$$\xrightarrow{y_i (w^T \phi(x_i) + b) \ge 1 + \xi_i} y_i (w^T \phi(x_i) + b) \ge 1$$



Both sides

KKT Explanation

$$\alpha_i[y_i(w^T\phi(x_i)+b)-1+\xi_i]=0$$
$$\beta_i\xi_i=0$$
$$0\le\alpha_i\le C$$



$if: 0<\alpha_i<C$

$then:$

$$\xrightarrow{\beta_i=C-\alpha_i \,\&\, \beta_i\xi_i=0} \xi_i=0$$

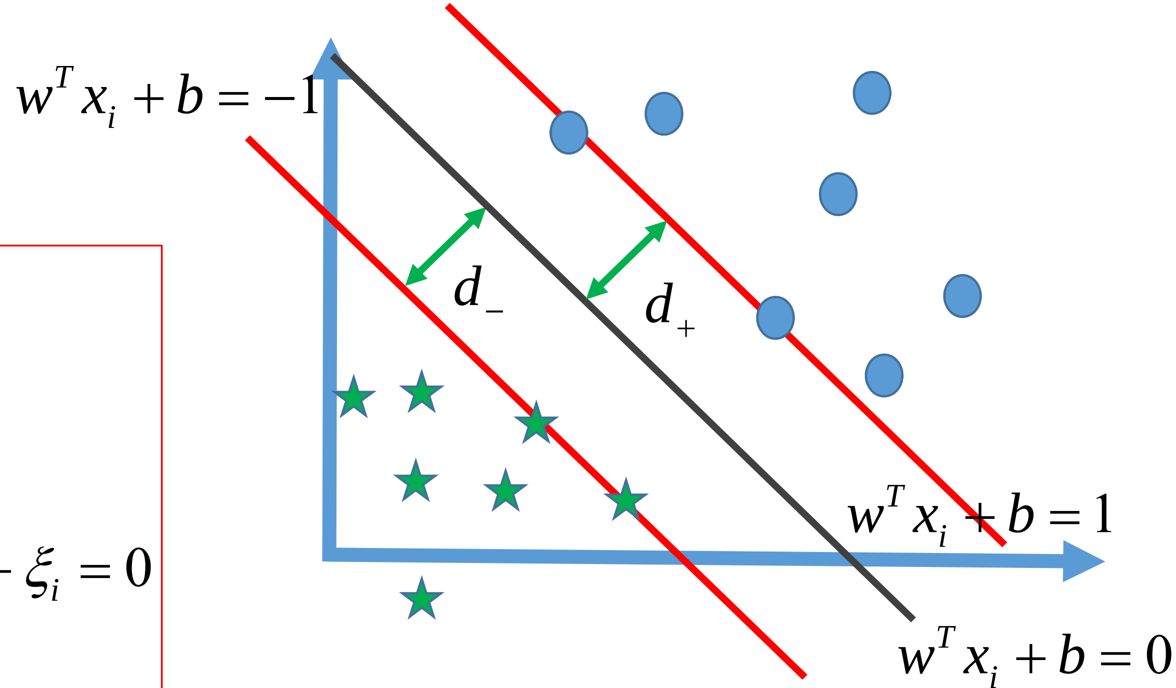
$$\xrightarrow{0<\alpha_i<C \,\&\, \alpha_i[y_i(w^T\phi(x_i)+b)-1+\xi_i]=0} y_i(w^T\phi(x_i)+b)-1+\xi_i=0$$

$$\longrightarrow y_i(w^T\phi(x_i)+b)=1$$

On Margin Edge

KKT Explanation

$$\alpha_i[y_i(w^T\phi(x_i)+b)-1+\xi_i]=0$$
$$\beta_i\xi_i=0$$
$$0\le\alpha_i\le C$$



$if:\alpha_i=C$

$then:$

$$\xrightarrow{\beta_i=C-\alpha_i\ \&\ \beta_i\xi_i=0}\xi_i\ge 0$$

$$\xrightarrow{\alpha_i[y_i(w^T\phi(x_i)+b)-1+\xi_i]=0}y_i(w^T\phi(x_i)+b)-1+\xi_i=0$$

$$\longrightarrow y_i(w^T\phi(x_i)+b)\le 1$$

On Margin Edge or Inner

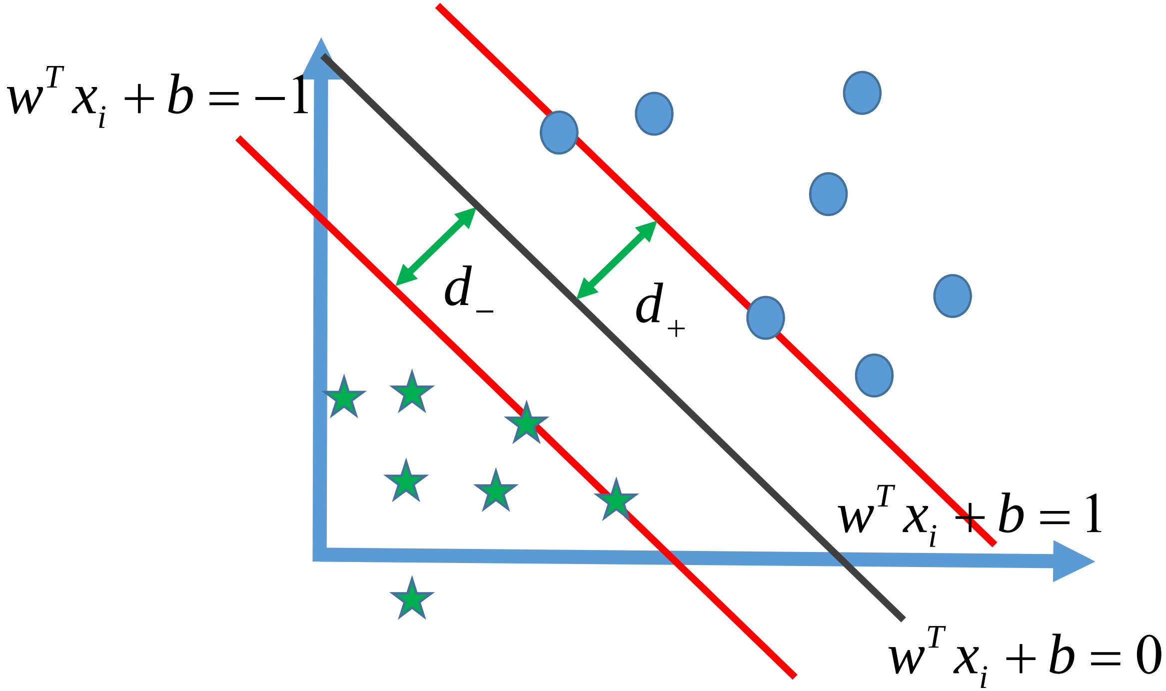
KKT Explanation: Analyzing cauchy Homework

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

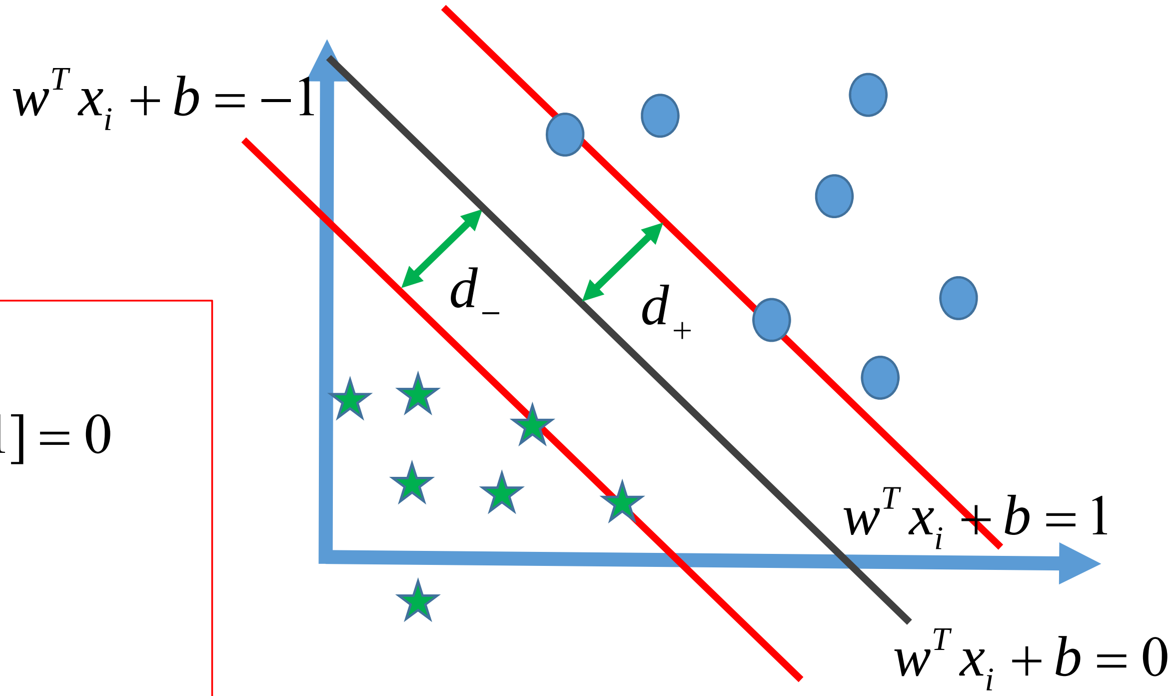
$$\beta_i\xi_i=0$$

$$\xi_i>0$$

$$\xi_i=0$$



How to solve b



$$if\ : 0 < \alpha_t < C$$

$$From\_KKT : \alpha_i[y_i(w^T\phi(x_i)+b)-1]=0$$

$$So : \phi(x_t)\_is\_on$$

$$(1).w^T\phi(x_t)+b=\pm 1$$

$$w^T\phi(x_t)+b=y_t$$

$$So : b = y_t - w^T\phi(x_t) = y_t - \sum_{i=1}^{l}\alpha_t y_i \phi(x_i)\phi(x_t)$$

$$= y_t - \sum_{i=1}^{l}\alpha_t y_i \kappa(x_i, x_t)$$