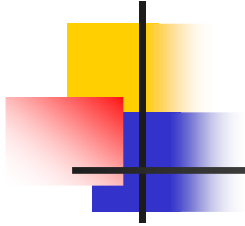




# Topics in Algorithms 2007

---

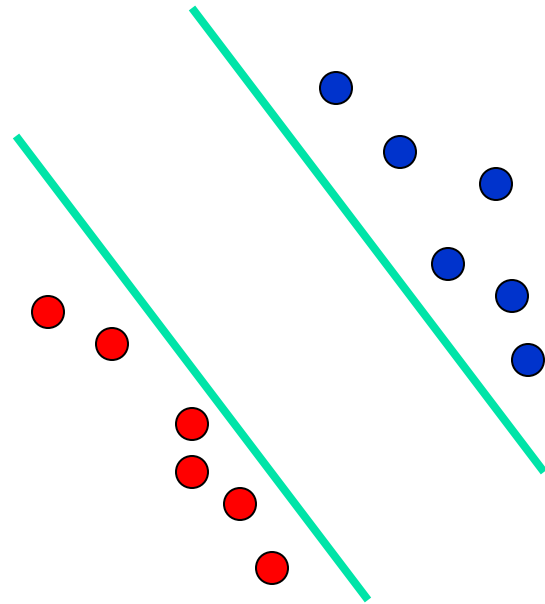
Ramesh Hariharan



# Support Vector Machines

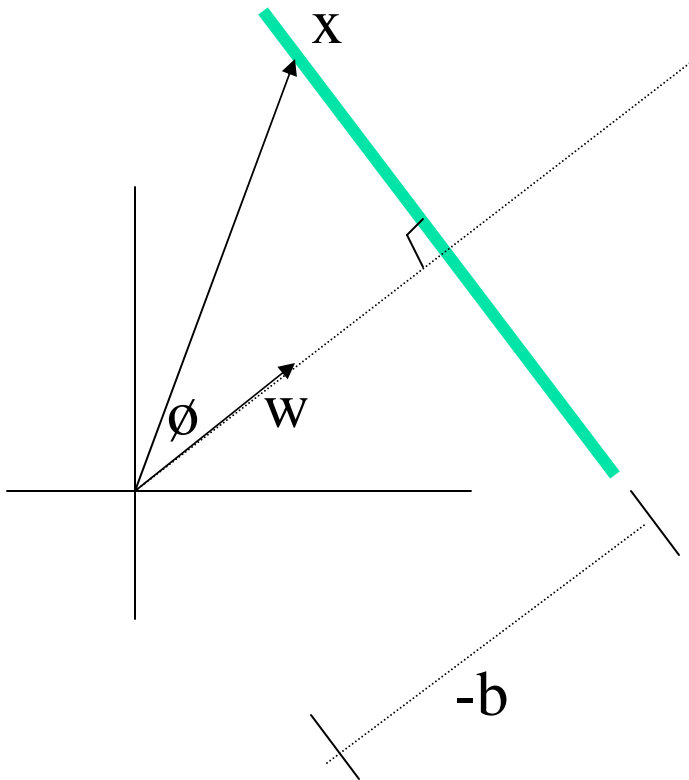
# Machine Learning

- How do learn *good* separators for 2 classes of points?
- Seperator could be linear or non-linear
- Maximize margin of separation



# Support Vector Machines

- Hyperplane  $w$



$$|w| = 1$$

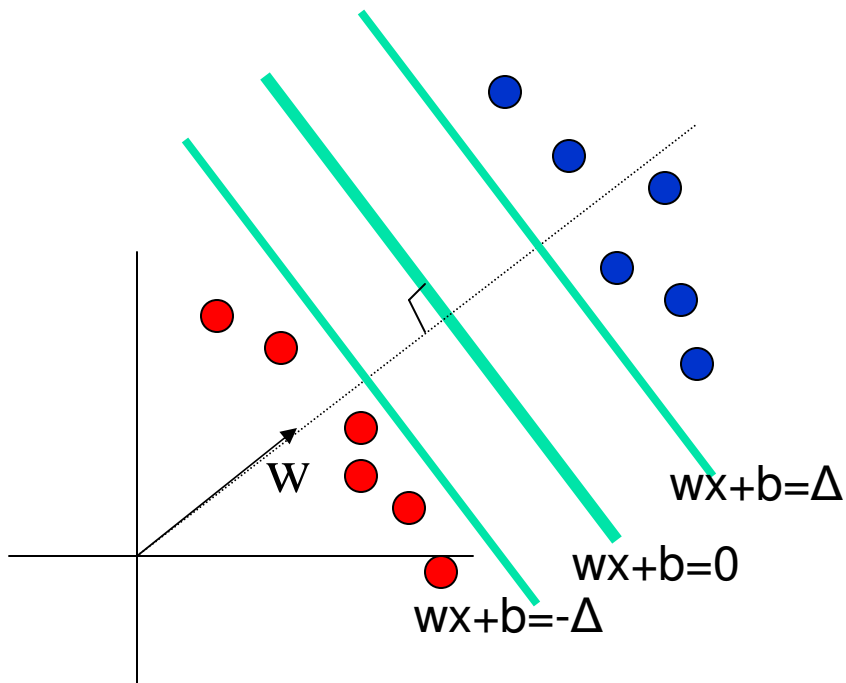
For all  $x$  on the hyperplane

$$w \cdot x = |w| |x| \cos(\phi) = |x| \cos(\phi) \\ = \text{constant} = -b$$

$$w \cdot x + b = 0$$

# Support Vector Machines

- Margin of separation

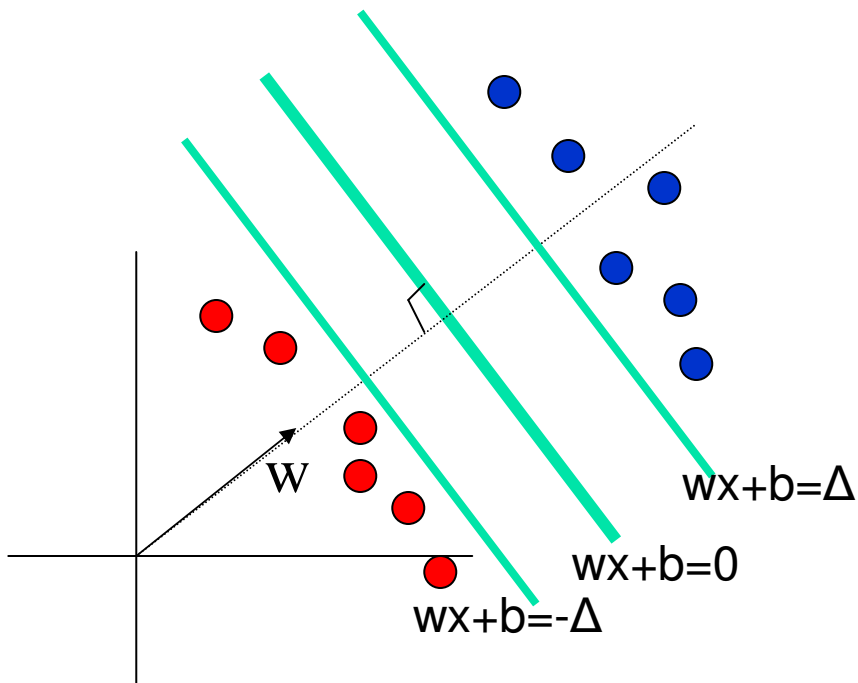


$|w| = 1$   
 $x \in \text{Blue}: wx+b \geq \Delta$   
 $x \in \text{Red}: wx+b \leq -\Delta$

maximize  $2 \Delta$   
 $w, b, \Delta$

# Support Vector Machines

- Eliminate  $\Delta$  by dividing by  $\Delta$



$$|w| = 1$$

$$x \in \text{Blue}: (w/\Delta) x + (b/\Delta) \geq 1$$

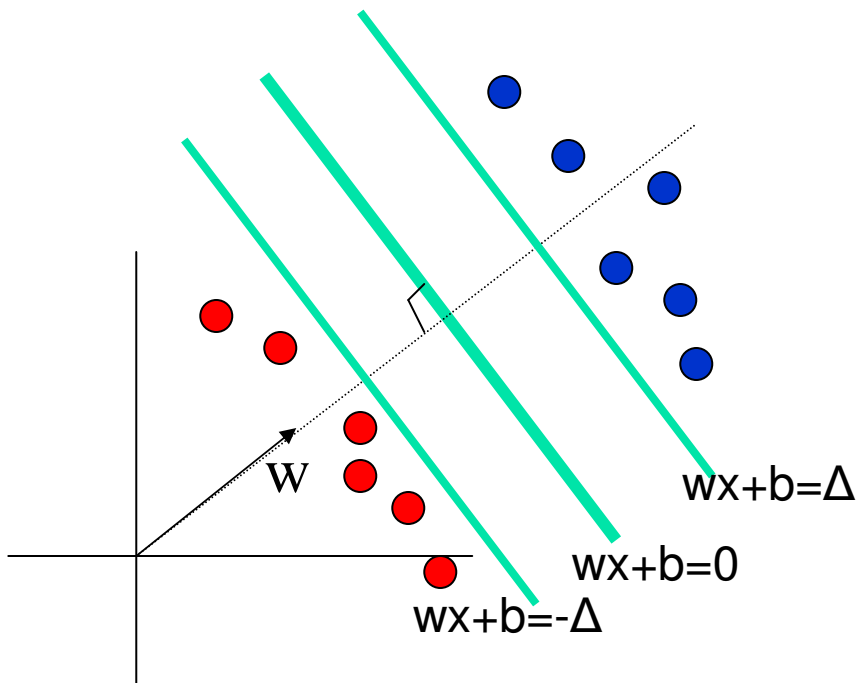
$$x \in \text{Red}: (w/\Delta) x + (b/\Delta) \leq -1$$

$$w' = w/\Delta \quad b' = b/\Delta$$

$$|w'| = |w|/\Delta = 1/\Delta$$

# Support Vector Machines

## ■ Perfect Separation Formulation



$x \in \text{Blue}: w'x+b' \geq 1$   
 $x \in \text{Red}: w'x+b' \leq -1$

minimize  $|w'|/2$   
 $w', b'$

minimize  $(w' \cdot w')/2$   
 $w', b'$



# Support Vector Machines

---

- Formulation allowing for misclassification

$x \in \text{Blue}: wx+b \geq 1$   
 $x \in \text{Red}: -(wx+b) \geq 1$

minimize  $(w.w)/2$   
 $w, b$

$x_i \in \text{Blue}: wx_i + b \geq 1 - \xi_i$   
 $x_i \in \text{Red}: -(wx_i + b) \geq 1 - \xi_i$

$\xi_i \geq 0$

minimize  $(w.w)/2 + C \sum \xi_i$   
 $w, b, \xi_i$



# Support Vector Machines

## ■ Duality

$$y_i (w x_i + b) + \xi_i \geq 1$$

$$\xi_i \geq 0$$

$y_i = \pm 1$ , class label

minimize  $(w \cdot w)/2 + C \sum \xi_i$   
 $w, b, \xi_i$

Primal

$$\sum \lambda_i y_i = 0$$

$$\lambda_i \geq 0$$

$$-\lambda_i \geq C$$

max  $\sum \lambda_i - ( \sum_i \sum_j \lambda_i \lambda_j y_i y_j (x_i \cdot x_j) )/2$   
 $\lambda_i$

Dual



# Support Vector Machines

- Duality (Primal  $\rightarrow$  Lagrangian  $\rightarrow$  Dual)
- If Primal is feasible then Primal=Lagrangian Primal

$$y_i (w x_i + b) + \xi_i \geq 1$$

$$\xi_i \geq 0$$

$y_i = \pm 1$ , class label

$$\min_{w, b, \xi_i} (w \cdot w) / 2 + C \sum \xi_i$$

Primal

=

$$\min_{w, b, \xi_i} \quad \max_{\lambda_i, \alpha_i \geq 0}$$

$$\begin{aligned} & (w \cdot w) / 2 + C \sum \xi_i \\ & - \sum_i \lambda_i (y_i (w x_i + b) + \xi_i - 1) \\ & - \sum_i \alpha_i (\xi_i - 0) \end{aligned}$$

Lagrangian Primal



# Support Vector Machines

- Lagrangian Primal  $\rightarrow$  Lagrangian Dual
- Lagrangian Primal  $\geq$  Lagrangian Dual

$$\min_{w, b, \xi_i} \max_{\lambda_i, \alpha_i \geq 0}$$

$$\frac{(w \cdot w)}{2} + C \sum \xi_i - \sum \lambda_i (y_i (w x_i + b) + \xi_i - 1) - \sum \alpha_i (\xi_i - 0)$$

Lagrangian Primal

$\geq$

$$\max_{\lambda_i, \alpha_i \geq 0} \min_{w, b, \xi_i}$$

$$\frac{(w \cdot w)}{2} + C \sum \xi_i - \sum \lambda_i (y_i (w x_i + b) + \xi_i - 1) - \sum \alpha_i (\xi_i - 0)$$

Lagrangian Dual



# Support Vector Machines

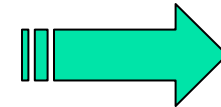
---

- Lagrangian Primal  $\geq$  Lagrangian Dual

- Proof

Consider a 2d matrix

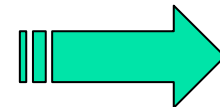
Find max in each row



LP

Find the smallest of these values

Find min in each column



LD

Find the largest of these values



# Support Vector Machines

- Can Lagrangian Primal = Lagrangian Dual ?

- Proof

Consider  $w^*$   $b^*$   $\xi^*$  optimal for primal

Find  $\lambda_i, \alpha_i \geq 0$  such that

minimizing over  $w, b, \xi$  gives  $w^*$

$b^*$   $\xi^*$

$$\sum_i \lambda_i (y_i (w^* x_i + b^*) + \xi_i^* - 1) = 0$$

$$\sum_i \alpha_i (\xi_i^* - 0) = 0$$

$$\begin{array}{ll} \max & \min \\ \lambda_i, \alpha_i \geq 0 & w, b, \xi_i \\ (w \cdot w) / 2 + C \sum \xi_i & \\ - \sum_i \lambda_i (y_i (w x_i + b) + \xi_i - 1) - \sum_i \alpha_i (\xi_i - 0) & \end{array}$$



# Support Vector Machines

- Can Lagrangian Primal = Lagrangian Dual ?

- Proof

Consider  $w^*$   $b^*$   $\xi_i^*$  optimal for primal

Find  $\lambda_i, \alpha_i \geq 0$  such that

$$\sum_i \lambda_i (y_i (w^* x_i + b^*) + \xi_i^* - 1) = 0$$

$$\sum_i \alpha_i (\xi_i^* - 0) = 0$$

$$\xi_i^* > 0 \text{ implies } \alpha_i = 0$$

$$y_i (w^* x_i + b^*) + \xi_i^* - 1 \neq 0 \text{ implies } \lambda_i = 0$$

$$\begin{aligned} & \max_{\lambda_i, \alpha_i \geq 0} && \min_{w, b, \xi_i} \\ & (w \cdot w) / 2 + C \sum \xi_i \\ & - \sum_i \lambda_i (y_i (w x_i + b) + \xi_i - 1) - \sum_i \alpha_i (\xi_i - 0) \end{aligned}$$



# Support Vector Machines

- Can Lagrangian Primal = Lagrangian Dual ?

- Proof

Consider  $w^*$   $b^*$   $\xi_i^*$  optimal for primal

Find  $\lambda_i, \alpha_i \geq 0$  such that

minimizing over  $w, b, \xi_i$  gives

$w^*, b^*, \xi_i^*$

at  $w^*, b^*, \xi_i^*$

$\delta / \delta w_j = 0, \delta / \delta \xi_i = 0, \delta / \delta b = 0$

and second derivatives should be non-neg at all places

$\max_{\lambda_i, \alpha_i \geq 0} \min_{w, b, \xi_i}$

$(w \cdot w) / 2 + C \sum \xi_i$

$-\sum_i \lambda_i (y_i (w x_i + b) + \xi_i - 1) - \sum_i \alpha_i (\xi_i - 0)$



# Support Vector Machines

- Can Lagrangian Primal = Lagrangian Dual ?

- Proof

Consider  $w^*$   $b^*$   $\xi_i^*$  optimal for primal

Find  $\lambda_i, \alpha_i \geq 0$  such that  
minimizing over  $w, b$  gives  $w^*, b^*$

$$w^* - \sum_i \lambda_i y_i x_i = 0$$

$$-\sum_i \lambda_i y_i = 0$$

$$-\lambda_i - \alpha_i + C = 0$$

second derivatives are always non-neg

$$\max_{\lambda_i, \alpha_i \geq 0} \quad \min_{w, b, \xi_i}$$

$$\begin{aligned} & (w \cdot w)/2 + C \sum \xi_i \\ & - \sum_i \lambda_i (y_i (w x_i + b) + \xi_i - 1) - \sum_i \alpha_i (\xi_i - 0) \end{aligned}$$



# Support Vector Machines

- Can Lagrangian Primal = Lagrangian Dual ?

- Proof

Consider  $w^*$   $b^*$   $\xi_i^*$  optimal for primal

Find  $\lambda_i, \alpha_i \geq 0$  such that

$$\xi_i^* > 0 \text{ implies } \alpha_i = 0$$

$$y_i (w^* x_i + b^*) + \xi_i^* - 1 \neq 0 \text{ implies } \lambda_i = 0$$

$$w^* - \sum_i \lambda_i y_i x_i = 0$$

$$-\sum_i \lambda_i y_i = 0$$

$$-\lambda_i - \alpha_i + C = 0$$

Such a  $\lambda_i, \alpha_i \geq 0$  always exists!!!!

$$\max_{\lambda_i, \alpha_i \geq 0} \quad \min_{w, b, \xi_i}$$

$$\begin{aligned} & (w \cdot w) / 2 + C \sum \xi_i \\ & - \sum_i \lambda_i (y_i (w x_i + b) + \xi_i - 1) - \sum_i \alpha_i (\xi_i - 0) \end{aligned}$$



# Support Vector Machines

---

- Proof that appropriate Lagrange Multipliers always exist?

- Roll all primal variables into  $w$   
lagrange multipliers into  $\lambda$

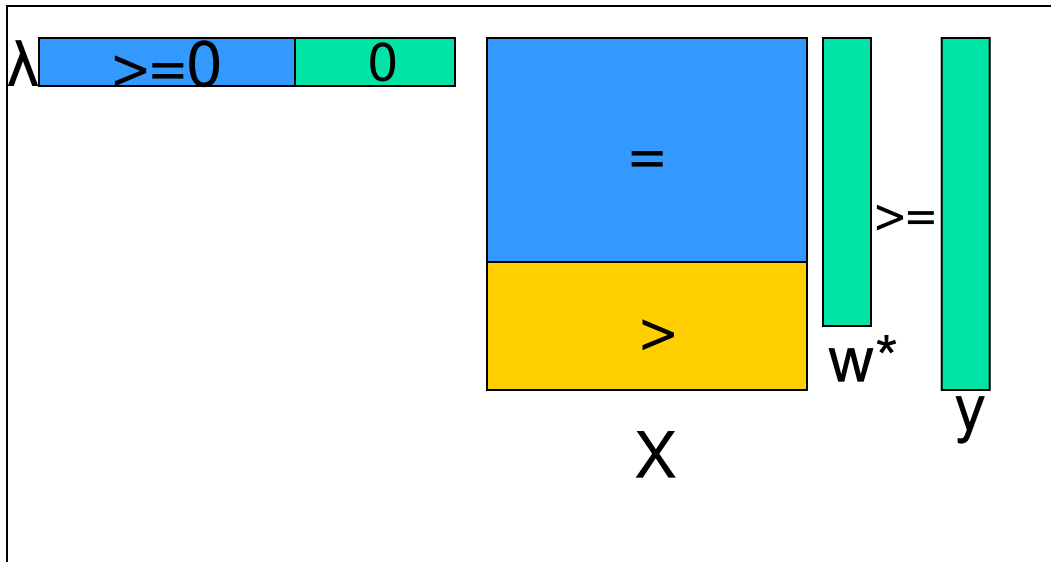
$$\begin{array}{l} \min f(w) \\ w \\ Xw \geq y \end{array}$$

$$\begin{array}{l} \min \max f(w) - \lambda (Xw - y) \\ w \quad \lambda \geq 0 \end{array}$$

$$\begin{array}{l} \max \min f(w) - \lambda (Xw - y) \\ \lambda \geq 0 \quad w \end{array}$$

# Support Vector Machines

- Proof that appropriate Lagrange Multipliers always exist?



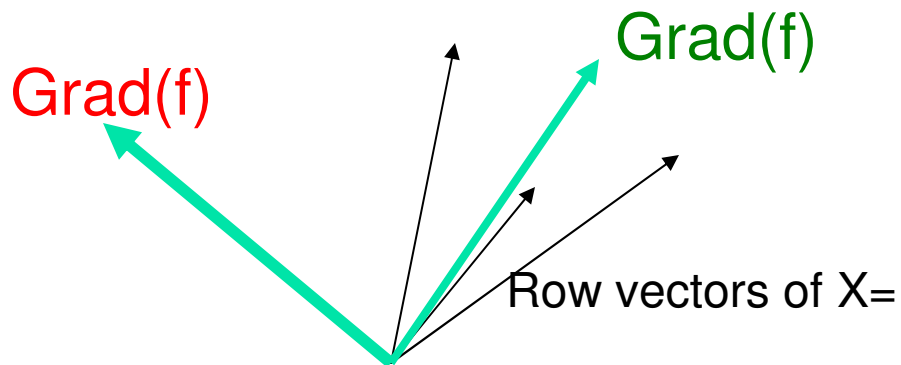
Claim: This is satisfiable

$\text{Grad}(f) \text{ at } w^* =$

$$\lambda \geq 0 \quad X =$$

# Support Vector Machines

- Proof that appropriate Lagrange Multipliers always exist?



Claim: This is satisfiable

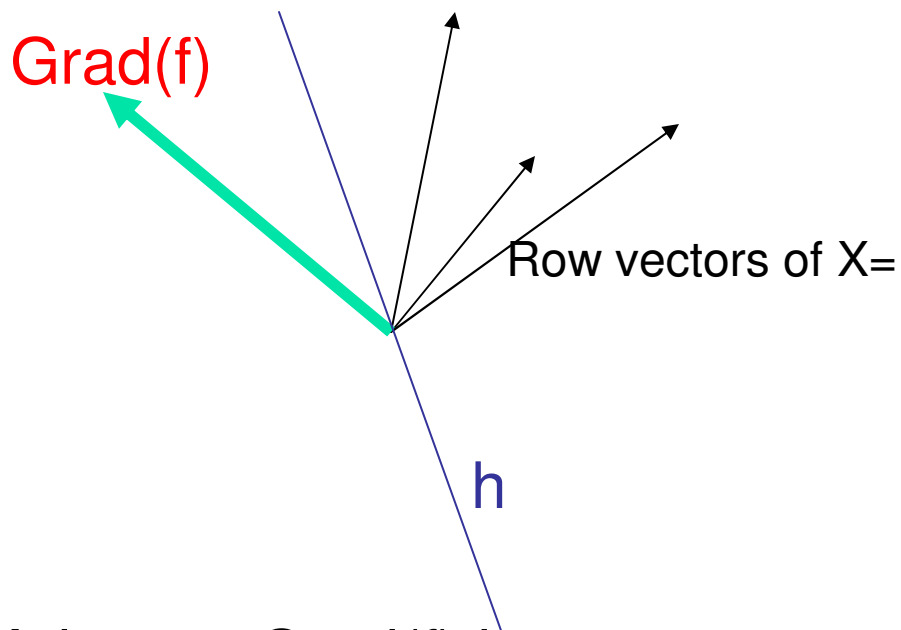
$$\text{Grad}(f) =$$

$$\lambda \geq 0$$

$$X=$$

# Support Vector Machines

- Proof that appropriate Lagrange Multipliers always exist?



$$X \cdot h \geq 0, \text{Grad}(f) \cdot h < 0$$

$w^* + h$  is feasible and  $f(w^* + h) < f(w^*)$  for small enough  $h$

Claim: This is satisfiable

$$\text{Grad}(f) =$$

$$\lambda \geq 0$$

$$X =$$



# Support Vector Machines

- Finally the Lagrange Dual

$$\begin{aligned} & \max_{\lambda_i, \alpha_i \geq 0} \min_{w, b, \xi_i} \\ & (w \cdot w) / 2 + C \sum \xi_i \\ & - \sum_i \lambda_i (y_i (w x_i + b) + \xi_i - 1) - \sum_i \alpha_i (\xi_i - 0) \end{aligned}$$

$$w - \sum_i \lambda_i y_i x_i = 0$$

$$- \sum_i \lambda_i y_i = 0$$

$$- \lambda_i - \alpha_i + C = 0$$

Rewrite in final dual form

$$\sum \lambda_i y_i = 0$$

$$\lambda_i \geq 0$$

$$- \lambda_i \geq -C$$

$$\max_{\lambda_i} \sum \lambda_i - ( \sum_i \sum_j \lambda_i \lambda_j y_i y_j (x_i \cdot x_j) ) / 2$$



# Support Vector Machines

## ■ Karush-Kuhn-Tucker conditions

$$\sum_i \lambda_i (y_i (w^* x_i + b^*) + \xi_i^* - 1) = 0$$

$$\sum_i \alpha_i (\xi_i^* - 0) = 0$$

$$-\lambda_i - \alpha_i + C = 0$$

$$\text{If } \xi_i^* > 0 \quad \alpha_i = 0 \quad \lambda_i = C$$

$$\text{If } y_i (w^* x_i + b^*) + \xi_i^* - 1 > 0$$

$$\lambda_i = 0$$

$$\xi_i^* = 0$$

$$\text{If } 0 < \lambda_i < C$$

$$y_i (w^* x_i + b^*) = 1$$

Rewrite in final dual form

$$\sum \lambda_i y_i = 0$$

$$\lambda_i \geq 0$$

$$-\lambda_i \geq C$$

$$\max_{\lambda_i} \sum \lambda_i - \left( \sum_i \sum_j \lambda_i \lambda_j y_i y_j (x_i \cdot x_j) \right) / 2$$