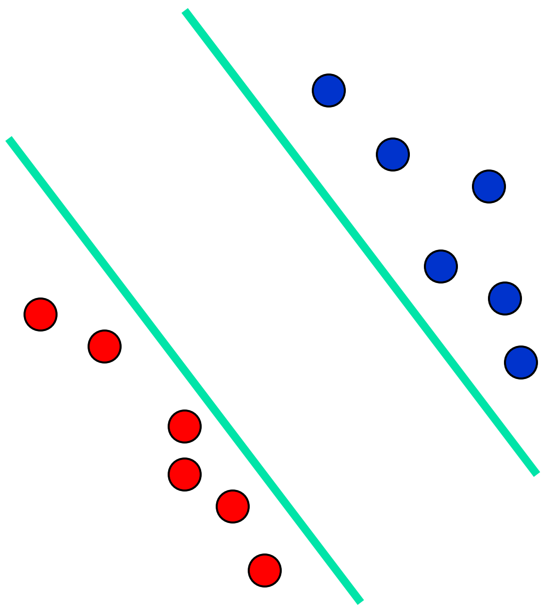
Topics in Algorithms 2007

Ramesh Hariharan

Support Vector Machines

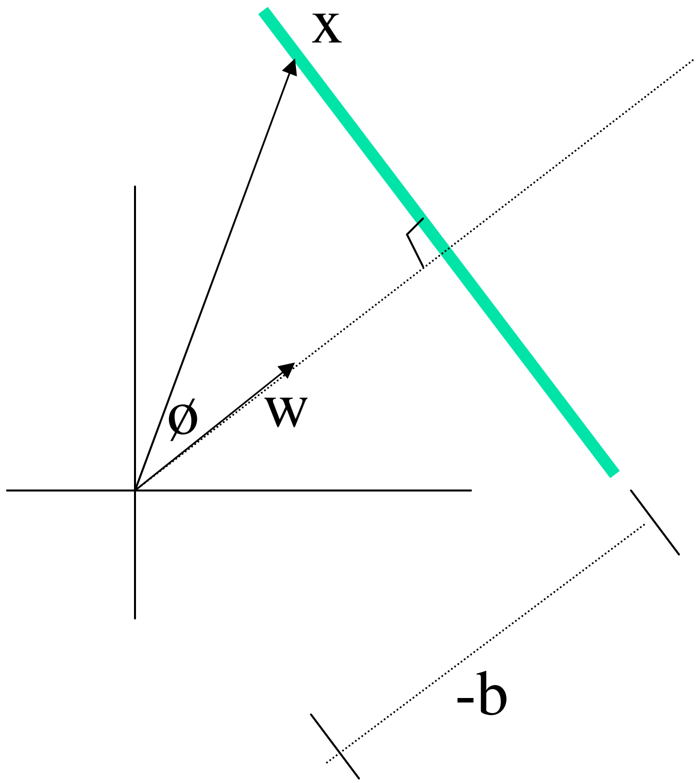
Machine Learning

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



Support Vector Machines

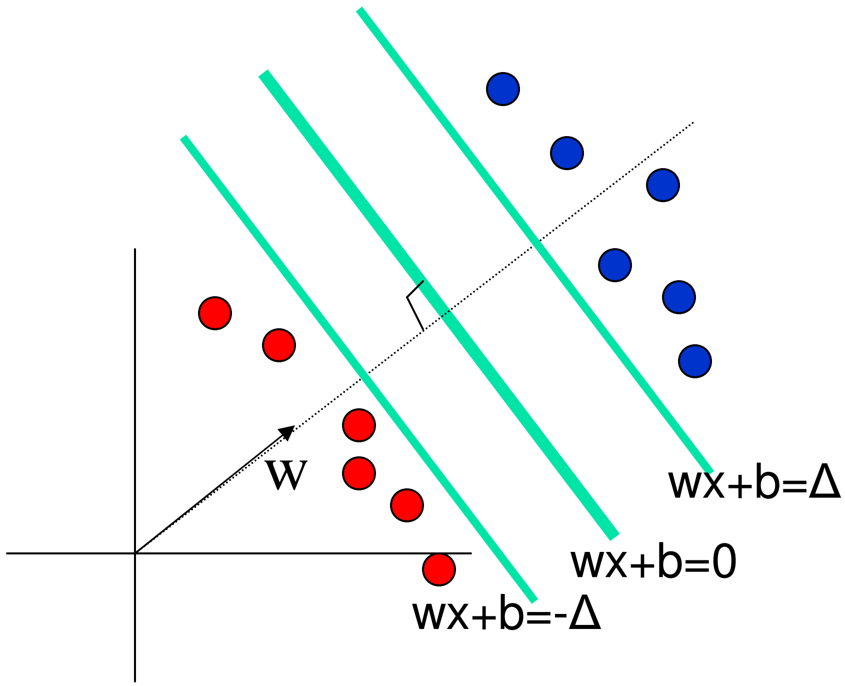
- Hyperplane w



|w| = 1
For all x on the hyperplane
w.x = |w||x| cos(ø)= |x|cos (ø)
= constant = -b
w.x+b=0

Support Vector Machines

- Margin of separation



$|w| = 1$

x Є Blue: $wx+b >= \Delta$

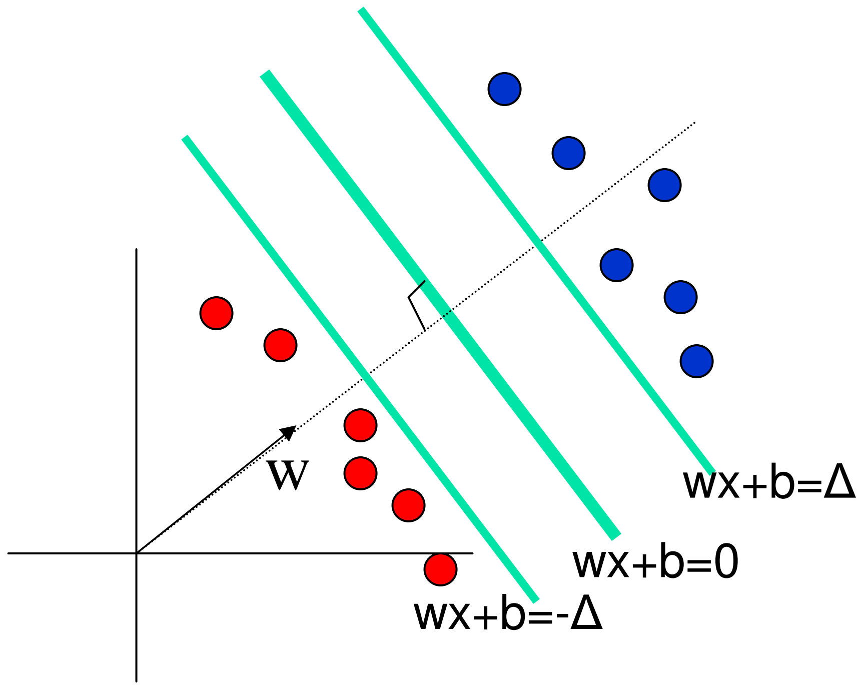
x Є Red: $wx+b <= -\Delta$

maximize $2\ \Delta$

w,b,Δ

Support Vector Machines

- Eliminate Δ by dividing by Δ



$$|w| = 1$$

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

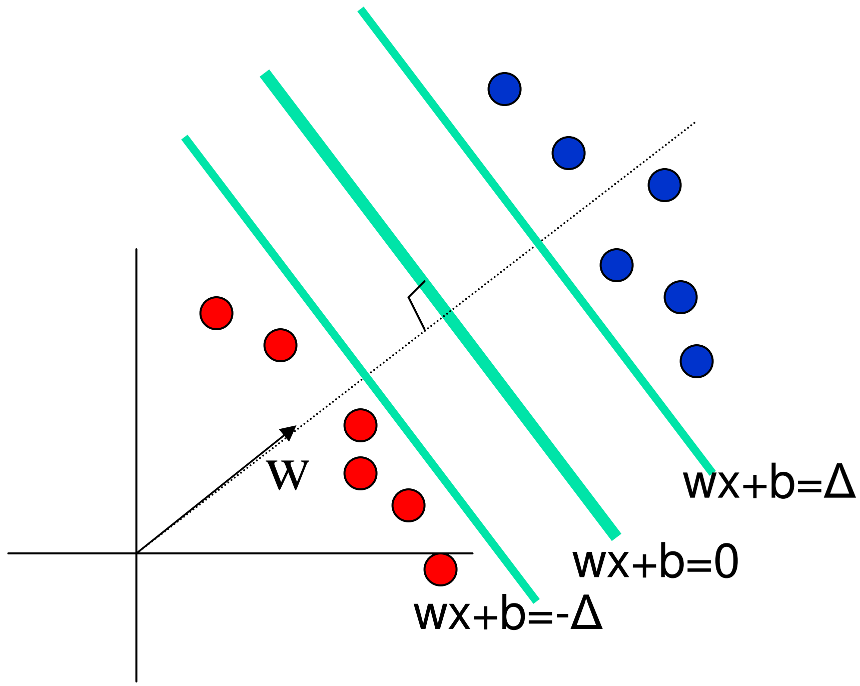
$$x \in \text{Red: } (w/\Delta)\, x + (b/\Delta) <= -1$$

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

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

Support Vector Machines

- Perfect Separation Formulation



x Є Blue: w’x+b’ >= 1
x Є Red: w’x+b’ <= -1

minimize |w’|/2
w’,b’

minimize (w’.w’)/2
w’,b’

Support Vector Machines

- Formulation allowing for misclassification

$x \in$ Blue: $wx+b >= 1$
$x \in$ Red: $-(wx+b) >= 1$

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

$x_i \in$ Blue: $wx_i + b >= 1-\xi_i$
$x_i \in$ Red: $-(wx_i + b) >= 1-\xi_i$

$\xi_i >= 0$

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

Support Vector Machines

- Duality

$y_i (wx_i + b) + \xi_i >= 1$

$\xi_i >= 0$

y_i=+/-1, class label

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

Primal

$\Sigma\ \lambda_i\ y_i\ = 0$

$\lambda_i >= 0$

$-\lambda_i >= C$

max $\Sigma\lambda_i - (\ \Sigma_i\Sigma_j\ \lambda_i\lambda_j y_i y_j\ (x_i.x_j)\ )/2$
λ_i

Dual

Support Vector Machines

- Duality (Primal → Lagrangian → Dual)
- If Primal is feasible then Primal=Lagrangian Primal

Primal		Lagrangian Primal
$y_i (wx_i + b) + \xi_i >= 1$ $\xi_i >= 0$ y_i=+/-1, class label $\min_{w,b,\xi_i} (w.w)/2 + C \Sigma\xi_i$	=	$\min_{w,b,\xi_i} \max_{\lambda_i, \alpha_i >=0}$ $(w.w)/2 + C \Sigma\xi_i$ $- \Sigma_i \lambda_i (y_i (wx_i + b) + \xi_i - 1)$ $- \Sigma_i \alpha_i (\xi_i - 0)$

Support Vector Machines

- Lagrangian Primal → Lagrangian Dual
- Langrangian Primal >= Lagrangian Dual

$$\min_{w,b,\xi_i} \max_{\lambda_i, \alpha_i >=0} (w.w)/2 + C \Sigma \xi_i - \Sigma_i \lambda_i (y_i (wx_i+b)+\xi_i - 1) - \Sigma_i \alpha_i (\xi_i - 0)$$

Lagrangian Primal

>=

$$\max_{\lambda_i, \alpha_i >=0} \min_{w,b,\xi_i} (w.w)/2 + C \Sigma \xi_i - \Sigma_i \lambda_i (y_i (wx_i+b)+\xi_i - 1) - \Sigma_i \alpha_i (\xi_i - 0)$$

Lagrangian Dual

Support Vector Machines

- Lagrangian Primal >= Lagrangian Dual

- Proof

  Consider a 2d matrix

  Find max in each row
  Find the smallest of these values → LP

  Find min in each column
  Find the largest of these values → LD

Support Vector Machines

- Can Lagrangian Primal = Lagrangian Dual ?

- Proof

Consider w^* b^* ξ^* optimal for primal

Find $\lambda_i, \alpha_i >= 0$ such that

minimizing over w,b,ξ gives w^* b^* ξ^*

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

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

$$\max_{\lambda_i, \alpha_i >= 0} \min_{w,b,\xi_i} (w.w)/2 + C\, \Sigma \xi_i - \Sigma_i \lambda_i (y_i (w x_i + b) + \xi_i - 1) - \Sigma_i \alpha_i (\xi_i - 0)$$

Support Vector Machines

- Can Lagrangian Primal = Lagrangian Dual ?

- Proof

Consider w^* b^* ξ_i^* optimal for primal

Find $\lambda_i, \alpha_i >= 0$ such that

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

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

$\xi_i^* > 0$ implies $\alpha_i = 0$

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

$$\max_{\lambda_i, \alpha_i >= 0} \; \min_{w,b,\xi_i}$$

$$(w.w)/2 + C \, \Sigma \xi_i - \Sigma_i \lambda_i (y_i (w x_i + b) + \xi_i - 1) - \Sigma_i \alpha_i (\xi_i - 0)$$

Support Vector Machines

- Can Lagrangian Primal = Lagrangian Dual ?

- Proof

Consider w^* b^* ξ_i^* optimal for primal

Find $\lambda_i, \alpha_i >= 0$ such that

minimizing over w, b, ξ_i gives w^*, b^*, ξ_i^*

at w^*, b^*, ξ_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

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

$$(w.w)/2 + C \Sigma \xi_i$$
$$-\Sigma_i \lambda_i (y_i (wx_i + b) + \xi_i - 1) - \Sigma_i \alpha_i (\xi_i - 0)$$

Support Vector Machines

- Can Lagrangian Primal = Lagrangian Dual ?

- Proof

Consider w^* b^* ξ_i^* optimal for primal

Find $\lambda_i, \alpha_i >= 0$ such that

minimizing over w,b gives w^*,b^*

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

$-\Sigma_i \lambda_i y_i = 0$

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

second derivatives are always non-neg

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

$$(w.w)/2 + C \Sigma \xi_i$$
$$-\Sigma_i \lambda_i (y_i (w x_i + b) + \xi_i - 1) - \Sigma_i \alpha_i (\xi_i - 0)$$

Support Vector Machines

- Can Lagrangian Primal = Lagrangian Dual ?

- Proof

Consider w^* b^* ξ_i^* optimal for primal

Find $\lambda_i, \alpha_i >= 0$ such that

$\xi_i^* > 0$ implies $\alpha_i = 0$

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

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

$-\Sigma_i \lambda_i y_i = 0$

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

Such a $\lambda_i, \alpha_i >= 0$ always exists!!!!!

$$\max_{\lambda_i, \alpha_i >= 0} \; \min_{w,b,\xi_i}$$

$$(w.w)/2 + C \; \Sigma \xi_i - \Sigma_i \lambda_i (y_i (w x_i + b) + \xi_i - 1) - \Sigma_i \alpha_i (\xi_i - 0)$$

Support Vector Machines

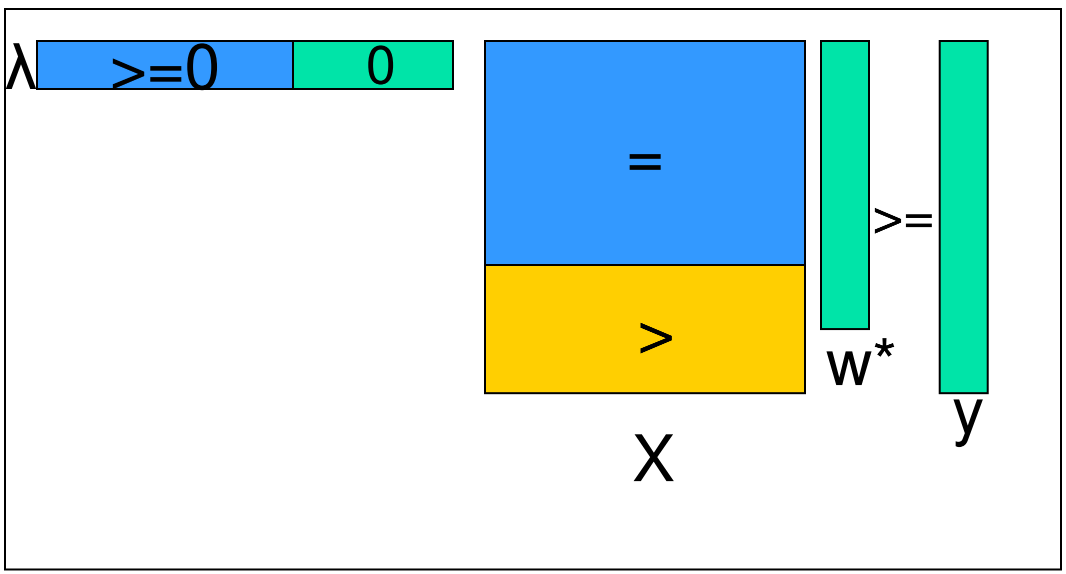
- Proof that appropriate Lagrange Multipliers always exist?

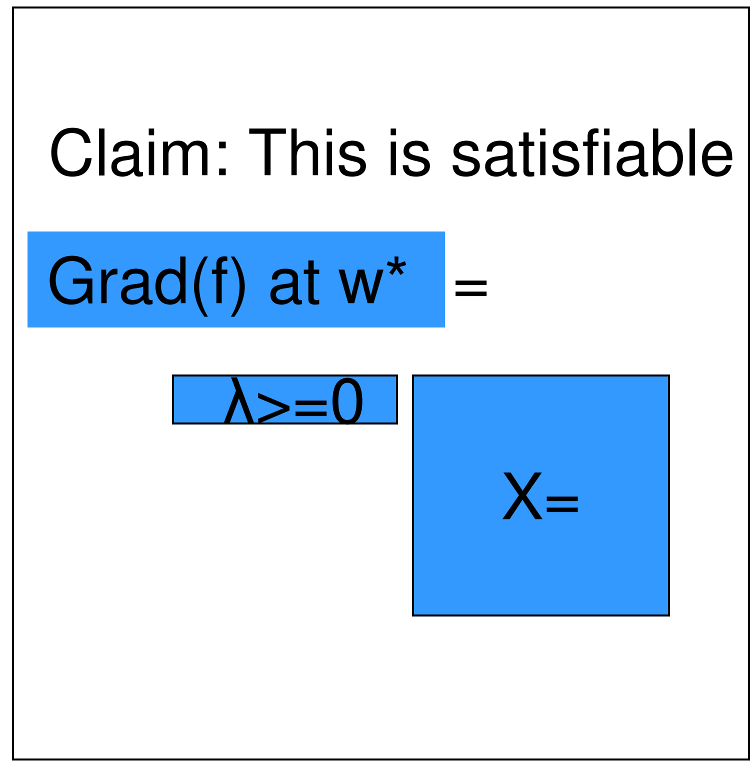
- Roll all primal variables into w
  lagrange multipliers into λ

$\min_{w} f(w)$ $Xw >= y$	$\min_{w} \max_{\lambda>=0} f(w) - \lambda (Xw-y)$	$\max_{\lambda>=0} \min_{w} f(w) - \lambda (Xw-y)$

Support Vector Machines

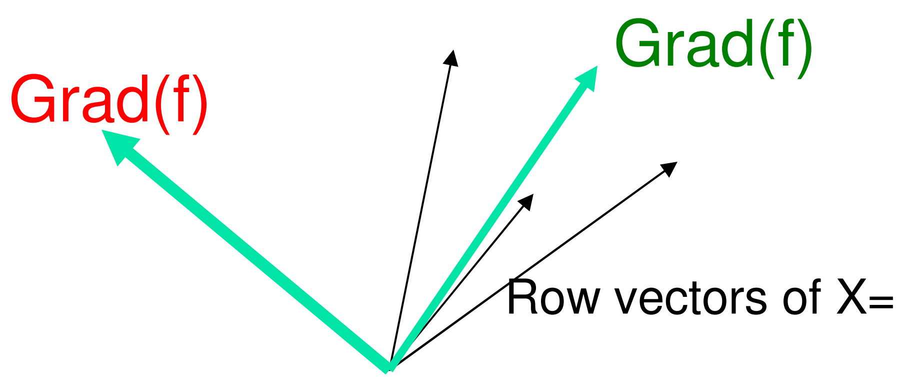
- Proof that appropriate Lagrange Multipliers always exist?





Support Vector Machines

- Proof that appropriate Lagrange Multipliers always exist?



Claim: This is satisfiable

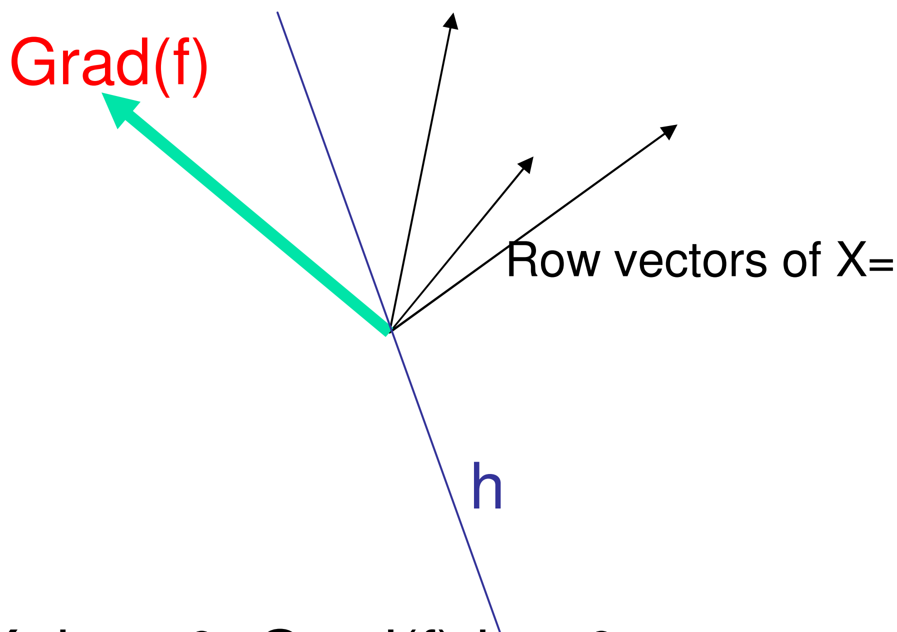
Grad(f) =

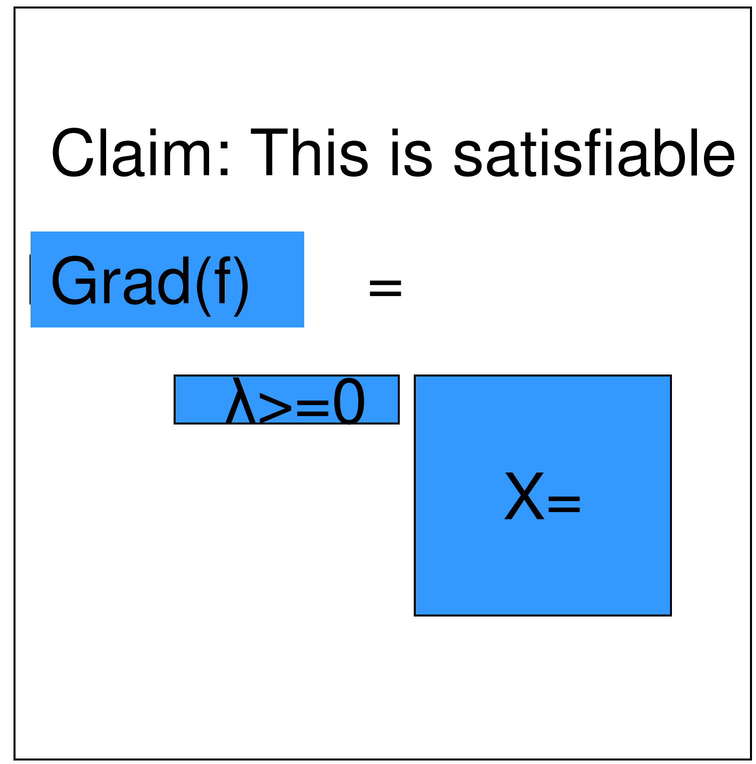
λ>=0

X=

Support Vector Machines

- Proof that appropriate Lagrange Multipliers always exist?





$X_=$ h >=0, Grad(f) h < 0

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

Support Vector Machines

- Finally the Lagrange Dual

$$\max_{\lambda_i, \alpha_i >= 0} \; \min_{w,b,\xi_i}$$

$$(w.w)/2 + C\,\Sigma\xi_i - \Sigma_i\lambda_i(y_i\,(wx_i + b) + \xi_i - 1) - \Sigma_i\alpha_i(\xi_i - 0)$$

$$w - \Sigma_i\lambda_i\, y_i\, x_i = 0$$

$$-\Sigma_i\lambda_i\, y_i = 0$$

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

Rewrite in final dual form

$$\Sigma\, \lambda_i\, y_i = 0$$

$$\lambda_i >= 0$$

$$-\lambda_i >= -C$$

$$\max_{\lambda_i} \Sigma\lambda_i - (\,\Sigma_i\Sigma_j\, \lambda_i\lambda_j y_i y_j\, (x_i.x_j)\,)/2$$

Support Vector Machines

- Karush-Kuhn-Tucker conditions

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

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

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

If $\xi_i^* > 0$ $\alpha_i = 0$ $\lambda_i = C$

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

$\lambda_i = 0$

$\xi_i^* = 0$

If $0 < \lambda_i < C$

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

Rewrite in final dual form

$\Sigma \lambda_i y_i = 0$

$\lambda_i >= 0$

$-\lambda_i >= C$

$\max_{\lambda_i} \Sigma \lambda_i - (\Sigma_i \Sigma_j \lambda_i \lambda_j y_i y_j (x_i . x_j))/2$