10701 Recitation 5
Duality and SVM

Ahmed Hefny

Outline

- Langrangian and Duality
  - The Lagrangian
  - Duality
  - Examples
- Support Vector Machines
  - Primal Formulation
  - Dual Formulation
  - Soft Margin and Hinge Loss

Lagrangian

- Consider the problem

$$\min_{x} f(x)$$

$$\text{s.t. } g_i(x) = 0$$

- Add a *Lagrange multiplier* for each constraint

$$L(x, u) = f(x) + \sum_i u_i g_i(x)$$

Lagrangian

- Lagrangian

$$L(x, u) = f(x) + \sum_i u_i g_i(x)$$

- Setting gradient to 0 gives
  - $g_i(x) = 0$ [Feasible point]
  - $\nabla f(x) + \sum_i u_i \nabla g_i(x) = 0$
    [Cannot decrease f except by violating constraints]

Lagrangian

- Consider the problem

$$\min_x f(x)$$

s.t. $$g_i(x) = 0$$

$$h_j(x) \leq 0$$

- Add a *Lagrange multiplier* for each constraint

$$L(x, u, \lambda) = f(x) + \sum_i u_i g_i(x) + \sum_j \lambda_j h_j(x)$$

Duality

Optimus Prime

Optimus Dual

Duality

- Primal problem

$$\min_x f(x)$$

$$\text{s.t.} \quad g_i(x) = 0$$

$$h_j(x) \leq 0$$

- Equivalent to

$$\min_x \max_{\lambda \geq 0, u} f(x) + \sum_i u_i g_i(x) + \sum_j \lambda_j h_j(x)$$

Duality

- Primal problem

$$\min_x f(x)$$

$$\text{s.t.} \quad g_i(x) = 0$$

$$h_j(x) \leq 0$$

- Equivalent to

$$\min_x \begin{cases} f(x) & x \text{ } is \text{ } feasible \\ \infty & o.w. \end{cases}$$

Duality

- Dual Problem

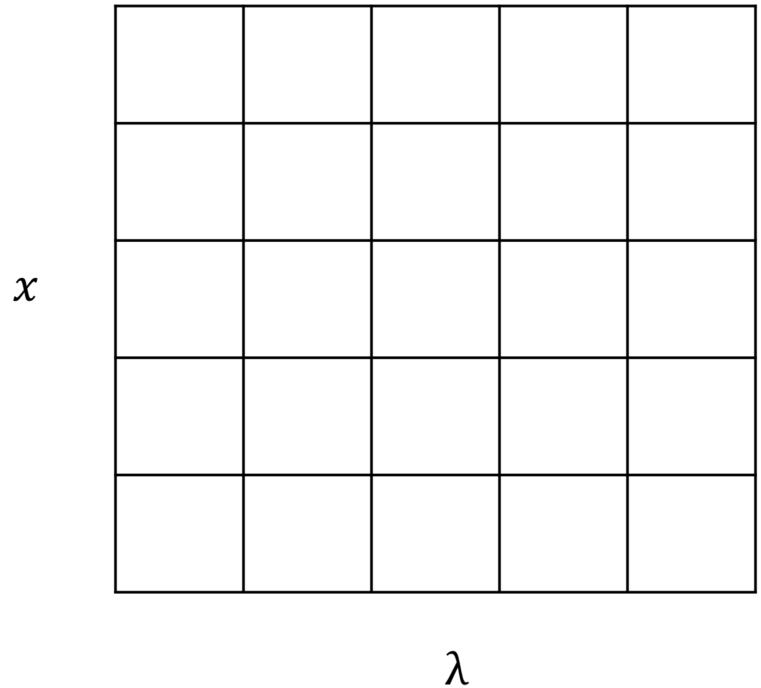
$$\max_{\lambda\geq 0,u} \underbrace{\min_{x} f(x) + \sum_i u_i g_i(x) + \sum_j \lambda_j h_j(x)}_{\text{Lagrangian Dual Function } L(\lambda,u)}$$

- Dual function:
  - Concave, regardless of the convexity of the primal
  - Lower bound on primal

Duality

Primal Problem

$$\min_{x} \max_{\lambda \geq 0} L(x, \lambda)$$

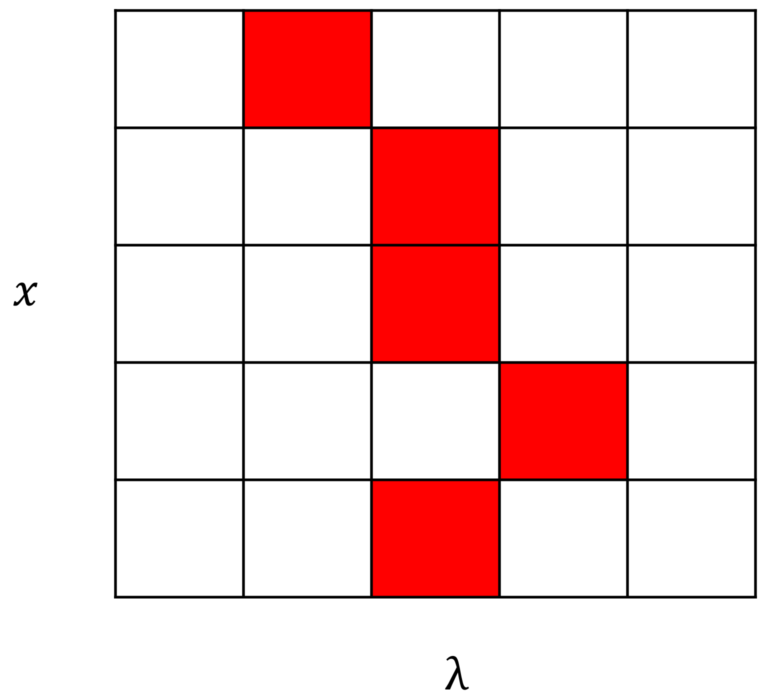


Duality

Primal Problem

$$\min_{x} \max_{\lambda \geq 0} L(x, \lambda)$$

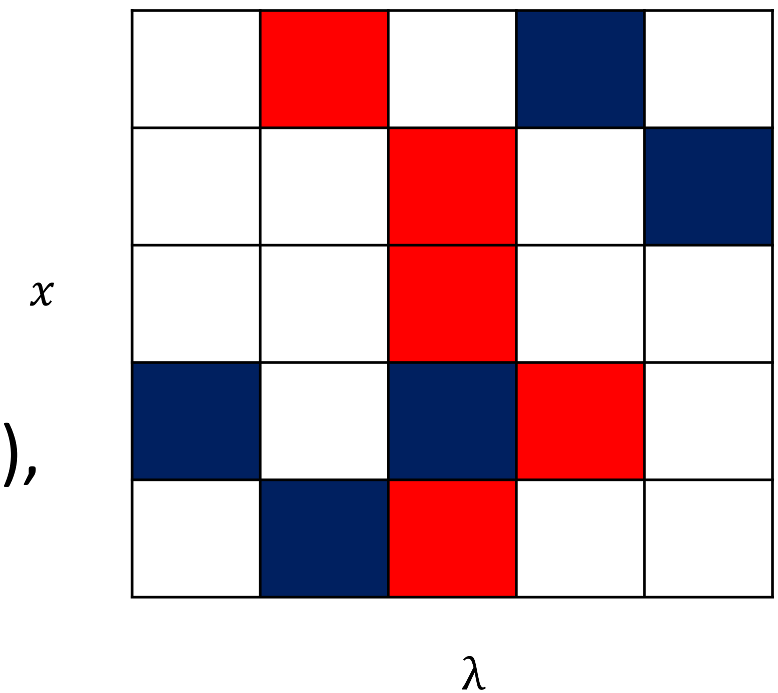
For each row (choice of x), pick the largest element then select the minimum.



Duality

Dual Problem

$$\max_{\lambda \geq 0} \min_{x} L(x, \lambda)$$

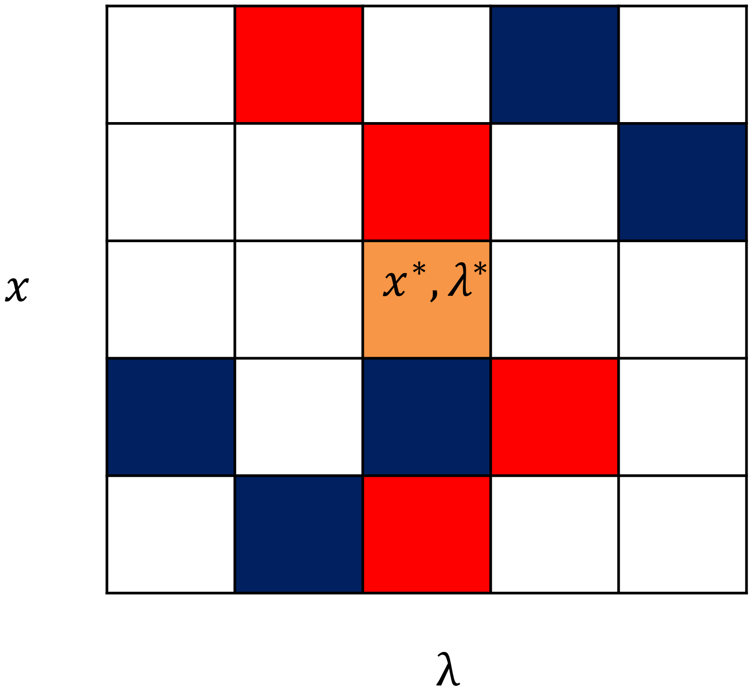


For each column (choice of λ), pick the smallest element then select the maximum.

Duality

Claim:

$$\min_{x} \max_{\lambda \geq 0} L(x,\lambda) \geq \max_{\lambda \geq 0} \min_{x} L(x,\lambda)$$



Duality

Claim:

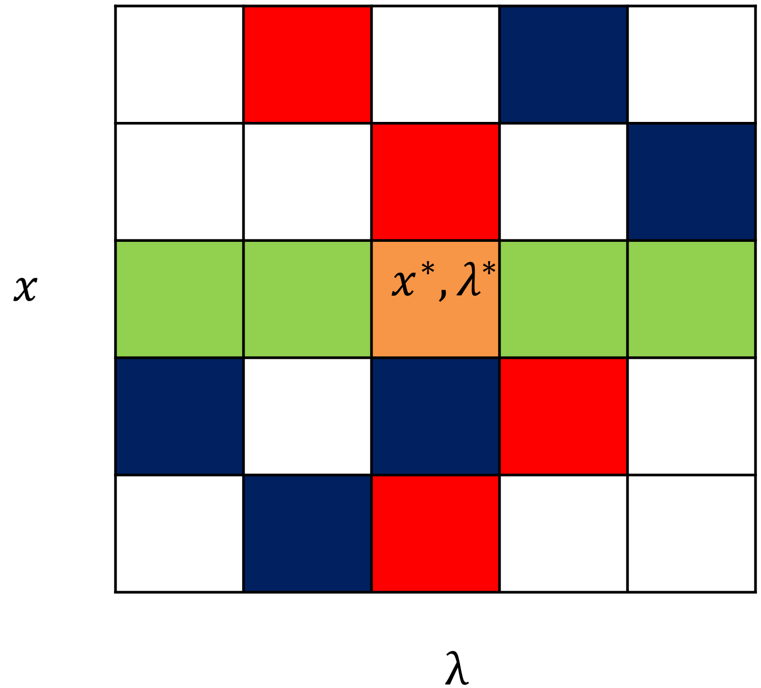
$$\min_{x} \max_{\lambda \geq 0} L(x, \lambda) \geq \max_{\lambda \geq 0} \min_{x} L(x, \lambda)$$

For any $\lambda \geq 0$

$$\min_{x} L(x, \lambda) \leq L(x^*, \lambda) \leq L(x^*, \lambda^*)$$

The difference between primal minimum
And dual maximum is called *duality gap*

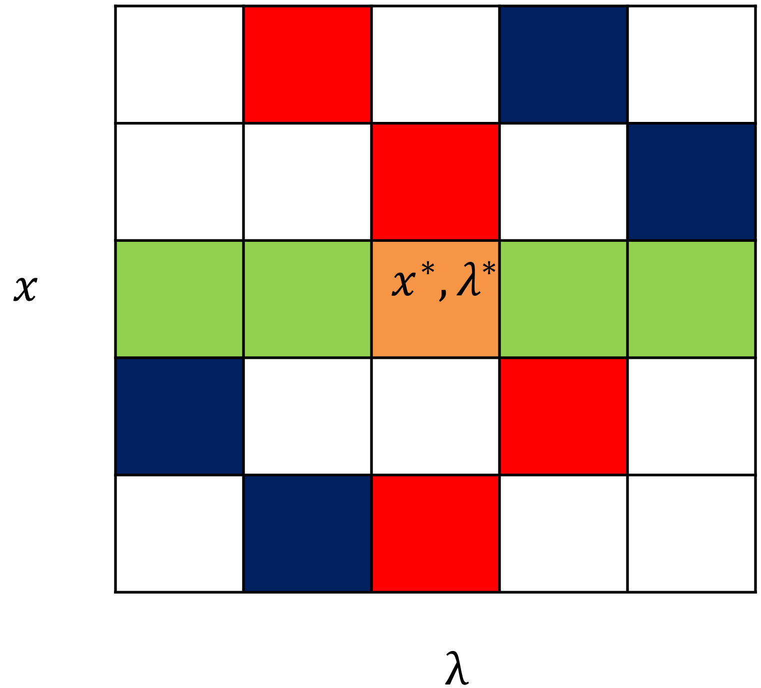
duality gap = 0 → Strong Duality



Duality

When does

$$\min_{x} \max_{\lambda \geq 0} L(x, \lambda) = \max_{\lambda \geq 0} \min_{x} L(x, \lambda)$$



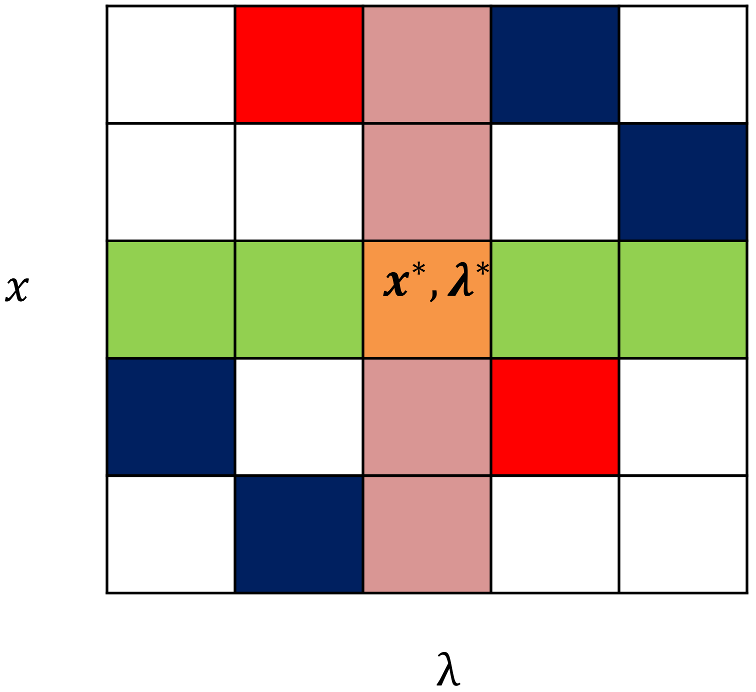
Duality

When does

$$\min_x \max_{\lambda \geq 0} L(x, \lambda) = \max_{\lambda \geq 0} \min_x L(x, \lambda)$$

x^*, λ^* is a saddle point

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)$$



Duality

When does

$$\min_{x} \max_{\lambda \geq 0} L(x, \lambda) = \max_{\lambda \geq 0} \min_{x} L(x, \lambda)$$

x^*, λ^* is a saddle point

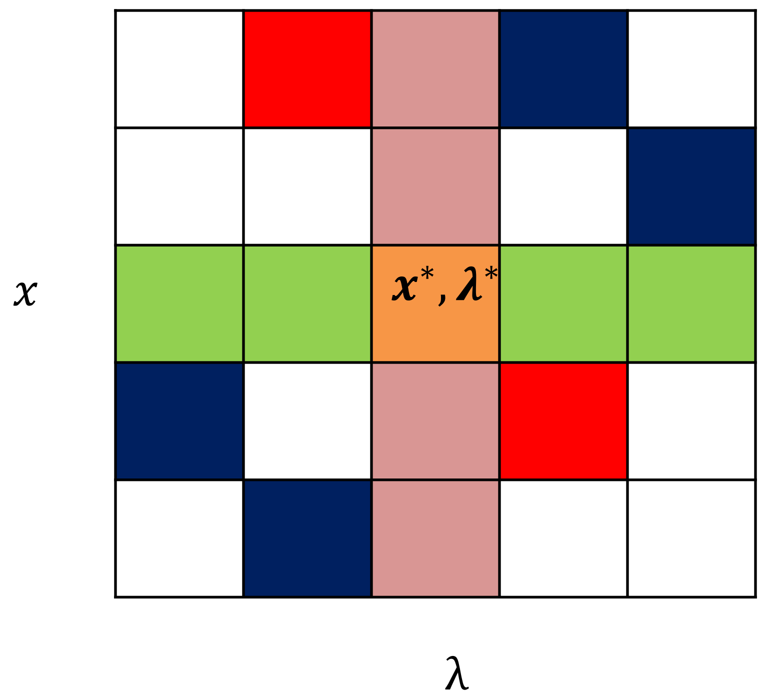
$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)$$

Necessity → By definition of dual

Sufficiency →

$$L(\lambda) = \min_{\mathrm{x}} L(x, \lambda) \leq L(x^*, \lambda^*)$$

$$L(\lambda^*) = L(x^*, \lambda^*)$$



Duality

When does

$$\min_x \max_{\lambda \geq 0} L(x,\lambda) = \max_{\lambda \geq 0} \min_x L(x,\lambda)$$

x^*, λ^* is a saddle point

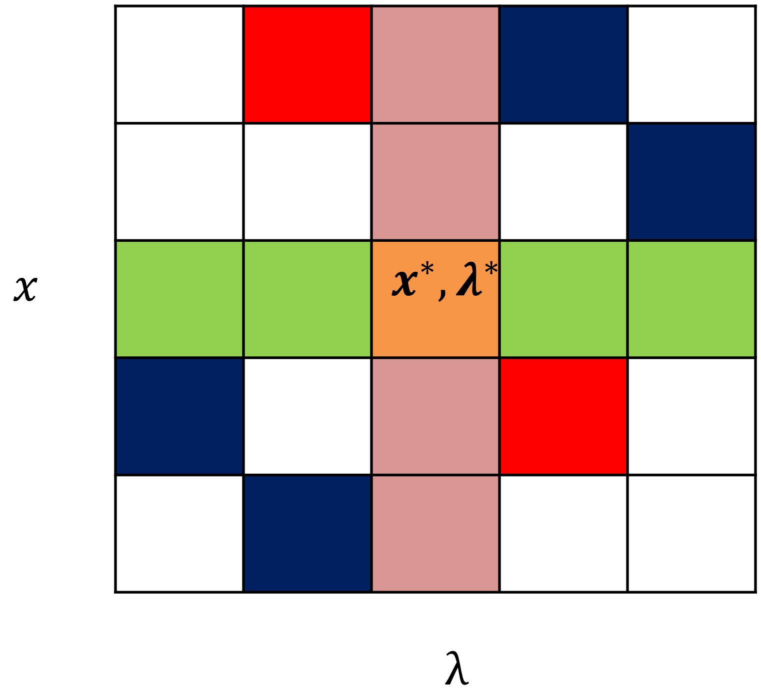
$$L(x^*,\lambda) \leq L(x^*,\lambda^*) \leq L(x,\lambda^*)$$

Necessity → By definition of dual

Sufficiency →

$$L(\lambda) = \min_x L(x,\lambda) \leq L(x^*,\lambda^*)$$

$$L(\lambda^*) = L(x^*,\lambda^*)$$



The dual at λ^* *is* the upper bound

Duality

- If strong duality holds, KKT conditions apply to optimal point
  - Stationary Point $\nabla L(x, u, \lambda) = 0$
  - Primal Feasibility
  - Dual Feasibility ($\lambda \geq 0$)
  - Complementary Slackness ($\lambda_i h_i(x) = 0$)
- KKT conditions are
  - Sufficient
  - Necessary under strong duality

Example: LP

- Primal

$$\min_x c^T x$$

$$\text{s.t. } Ax \geq b$$

Example: LP

- Primal

$$\min_x c^T x$$
$$\text{s.t. } Ax \geq b$$

- Lagrangian

$$L(x, \lambda) = c^T x - \lambda^T (Ax - b)$$

Example: LP

- Dual Function

$$L(\lambda) = \min_x c^T x - \lambda^T (Ax - b)$$

Example: LP

- Dual Function

$$L(\lambda) = \min_x c^T x - \lambda^T (Ax - b)$$

- Set gradient w.r.t x to 0

$$c \quad - A^T \lambda = 0$$

Example: LP

- Dual Function

$$L(\lambda) = \min_x c^T x - \lambda^T (Ax - b)$$

- Set gradient w.r.t x to 0

$$c - A^T \lambda = 0$$

- Dual Problem

$$\max_{\lambda \geq 0} \lambda^T b$$

s.t. $c - A^T \lambda = 0$

Why keep this as a constraint ?

Example: LASSO

- We will use duality to transform LASSO into a QP

Example: LASSO

Primal

$$\min \frac{1}{2} \|y - Xw\|^2 + \gamma \|w\|_1$$

What is the dual function in this case ?

Example: LASSO

Reformulated Primal

$$\min \frac{1}{2}\|y - z\|^2 + \gamma\|w\|_1$$

$$s.t. z = Xw$$

Dual

$$L(\lambda) = \min_{z,w} \frac{1}{2}\|y - z\|^2 + \gamma\|w\|_1 + \lambda^T(z - Xw)$$

Example: LASSO

Dual

$$L(\lambda) = \min_{z,w} \frac{1}{2} \|y - z\|^2 + \gamma \|w\|_1 + \lambda^T (z - Xw)$$

Setting gradient to zero gives

$$z = y - \lambda$$

$$\|X^T \lambda\|_\infty \leq \gamma$$

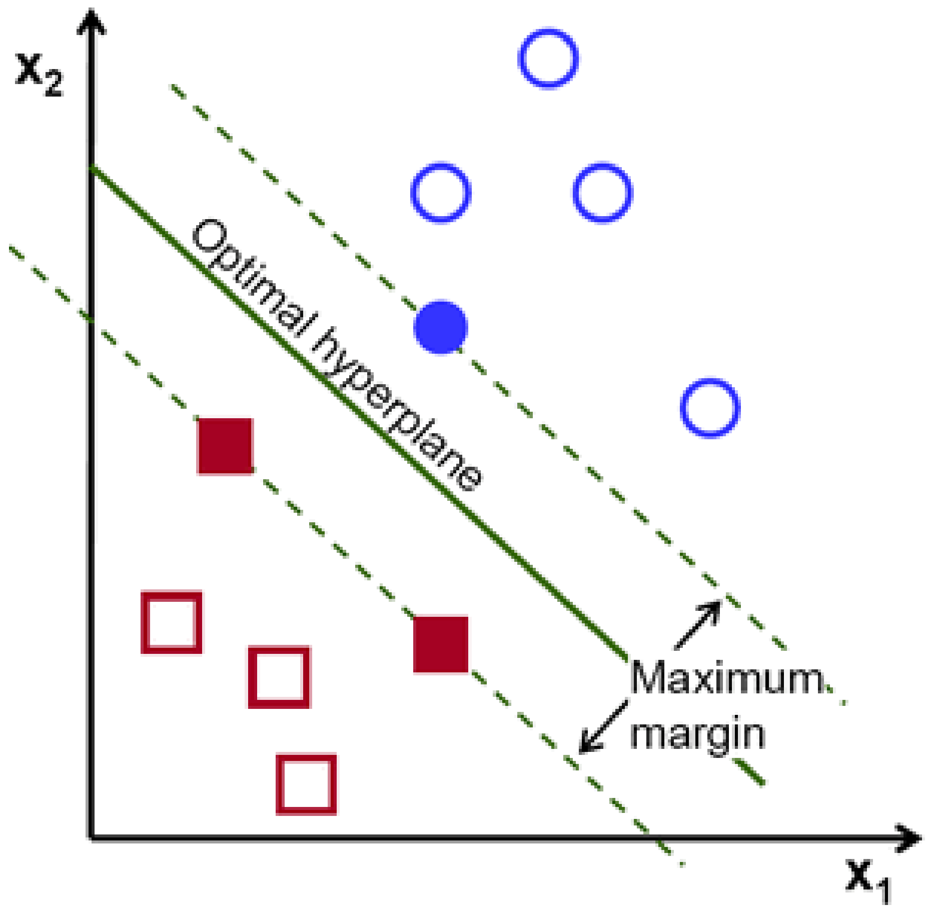
Example: LASSO

- Dual Problem

$$\max -\frac{1}{2}\|\lambda\|^2 + \lambda^T y$$

$$\text{s.t. } \|X^T\lambda\|_\infty \leq \gamma$$

Support Vector Machines



docs.opencv.org

Support Vector Machines

- Find the maximum margin hyper-plane
- "Distance" from a point x to the hyper-plane $\langle w, x_i \rangle + b = 0$ is given by
$$d_i = (\langle w, x_i \rangle + b)/\|w\|$$
- $Margin = \min_i y_i d_i = \frac{1}{\|w\|} \min_i (\langle w, x_i \rangle + b) y_i$
- Max Margin: $\max_{w,b} \frac{1}{\|w\|} \min_i (\langle w, x_i \rangle + b) y_i$

Support Vector Machines

- Max Margin

$$\max_{w,b} \frac{1}{\|w\|} \min_i (\langle w, x_i \rangle + b) y_i$$

- Unpleasant (max min ?)
- No Unique Solution

Support Vector Machines

- Max Margin

$$\max_{w,b} \frac{1}{\|w\|} \min_i (\langle w, x_i \rangle + b) y_i$$

s.t. ???

Support Vector Machines

- Max Margin

$$\max_{w,b} \frac{1}{\|w\|} \min_i (\langle w, x_i \rangle + b) y_i$$

$$\text{s.t. } \min_i (\langle w, x_i \rangle + b) y_i = 1$$

Support Vector Machines

- Max Margin

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

$$\text{s.t. } \min_i (\langle w, x_i \rangle + b) y_i = 1$$

Support Vector Machines

- Max Margin (Canonical Representation)

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

$$\text{s.t. } (\langle w, x_i \rangle + b) y_i \geq 1, \forall i$$

- QP, much better than

$$\max_{w,b} \frac{1}{\|w\|} \min_i (\langle w, x_i \rangle + b) y_i$$

SVM Dual Problem

Recall that the Lagrangian is formed by adding a Lagrange multiplier for each constraint.

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_i \alpha_i \left[(\langle w, x_i \rangle + b) y_i - 1\right]$$

SVM Dual Problem

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_i \alpha_i \left[(\langle w, x_i \rangle + b)y_i - 1\right]$$

Fix α and minimize w.r.t w, b:

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

$$\sum_i \alpha_i y_i = 0$$

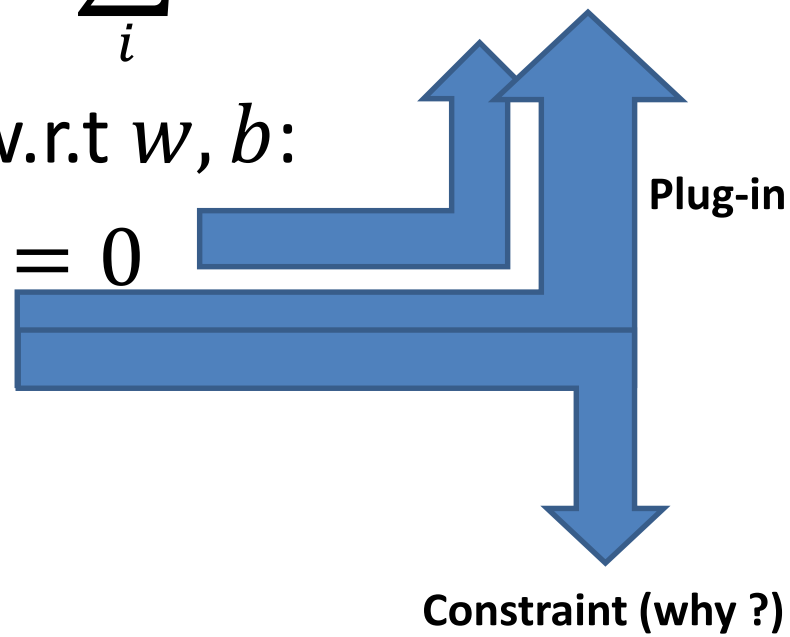
SVM Dual Problem

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_i \alpha_i \left[(\langle w, x_i\rangle + b)y_i - 1\right]$$

Fix α and minimize w.r.t w, b:

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

$$\sum_i \alpha_i y_i = 0$$



SVM Dual Problem

Dual Problem

$$\max -\frac{1}{2}\sum_{i}\sum_{j}\alpha_i\alpha_j y_i y_j \langle x_i, x_j\rangle + \sum_{i}\alpha_i$$

$$s.t. \sum_i \alpha_i y_i = 0$$

$$\alpha_i \geq 0$$

Another QP. So what ?

SVM Dual Problem

- Only Inner products → Kernel Trick
- Complementary Slackness → Support Vectors
- KKT conditions lead to Efficient optimization algorithms (compared to general QP solver)

SVM Dual Problem

- Classification of a test point

$$f(x) = \langle w, x \rangle + b = \sum_i \alpha_i y_i \langle x_i, x \rangle + b$$

- To get b use the fact that $y_i f(x_i) = 1$ for any support vector.
- For numerical stability, average over all support vectors.

Soft Margin SVM

Hard Margin SVM

$$\min_{\mathrm{w,b}} \sum_i E_\infty[1- \ (\langle w, x_i \rangle + b) y_i] + \frac{1}{2} \|w\|^2$$

, where

$$E_\infty[x] = \begin{cases} \infty & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Soft Margin SVM

Hard Margin SVM

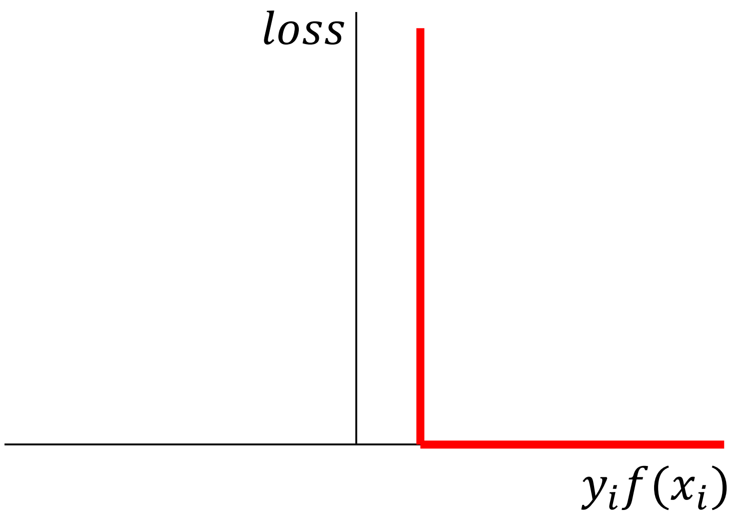
$$\min_{w,b} \sum_i E_\infty[1 - (\langle w, x_i \rangle + b) y_i] + \frac{1}{2} \|w\|^2$$

loss

regularization

, where

$$E_\infty[x] = \begin{cases} \infty & x \geq 0 \\ 0 & x < 0 \end{cases}$$



Soft Margin SVM

Relax it a little bit

$$\min_{\mathrm{w,b}} \sum_i E_C[1 - (\langle w, x_i \rangle + b) y_i] + \frac{1}{2} \|w\|^2$$

, where

$$E_C[x] = \begin{cases} Cx & x \geq 0 \\ 0 & x < 0 \end{cases}$$

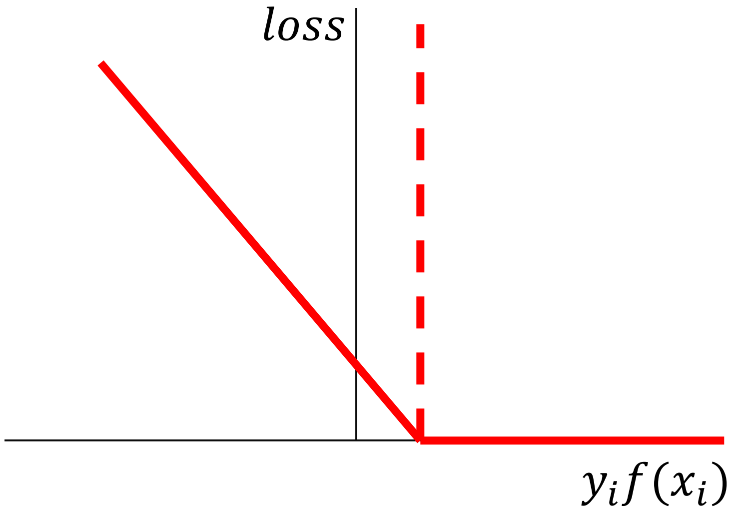
Soft Margin SVM

Relax it a little bit

$$\min_{\mathrm{w,b}} \sum_i E_C[1 - (\langle w, x_i \rangle + b)y_i] + \frac{1}{2}\|w\|^2$$

, where

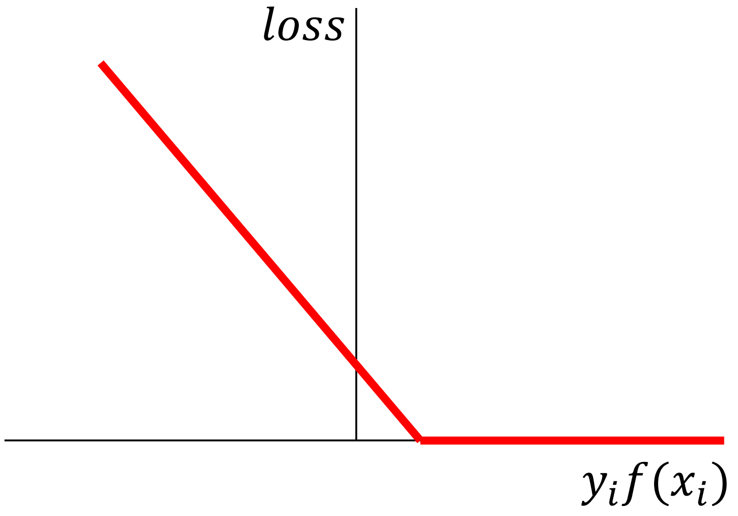
$$E_C[x] = \begin{cases} Cx & x \geq 0 \\ 0 & x < 0 \end{cases}$$



Soft Margin SVM

Relax it a little bit

$$\min_{\mathrm{w,b}} \; C\sum_i [1-(\langle w, x_i\rangle + b)y_i]_+ + \frac{1}{2}\|w\|^2$$



Soft Margin SVM

Equivalent Formulation

$$\min_{\mathrm{w,b},\zeta} \; C\sum_i \zeta_i + \frac{1}{2}\|w\|^2$$

s.t. $\zeta_i \geq 0$

$$(\langle w, x_i \rangle + b)y_i \geq 1 - \zeta_i$$

Conclusions

- Duality allows for establishing a lower bound on minimization problem.
- Key idea
  - "min max" upper bounds "max min"
- Strong Duality → Necessity of KKT Conditions
- Duality on SVMs
  - Kernel Trick
  - Support Vectors
- Soft Margin SVM = Hinge Loss

Resources

- Bishop, “Pattern Recognition and Machine Learning”, Chp 7
- Gordon & Tibshirani, 10725 Optimization (Fall 2012) Lecture Slides: http://www.cs.cmu.edu/~ggordon/10725-F12/schedule.html
- Fiterau, Kernels and SVM “http://alex.smola.org/teaching/cmu2013-10-701/slides/6_Recitation_Kernels.pdf”