

UVA CS : Machine Learning

Lecture Extra: Support Vector Machine Optimization with Dual

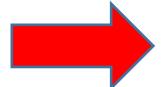
Dr. Yanjun Qi

University of Virginia

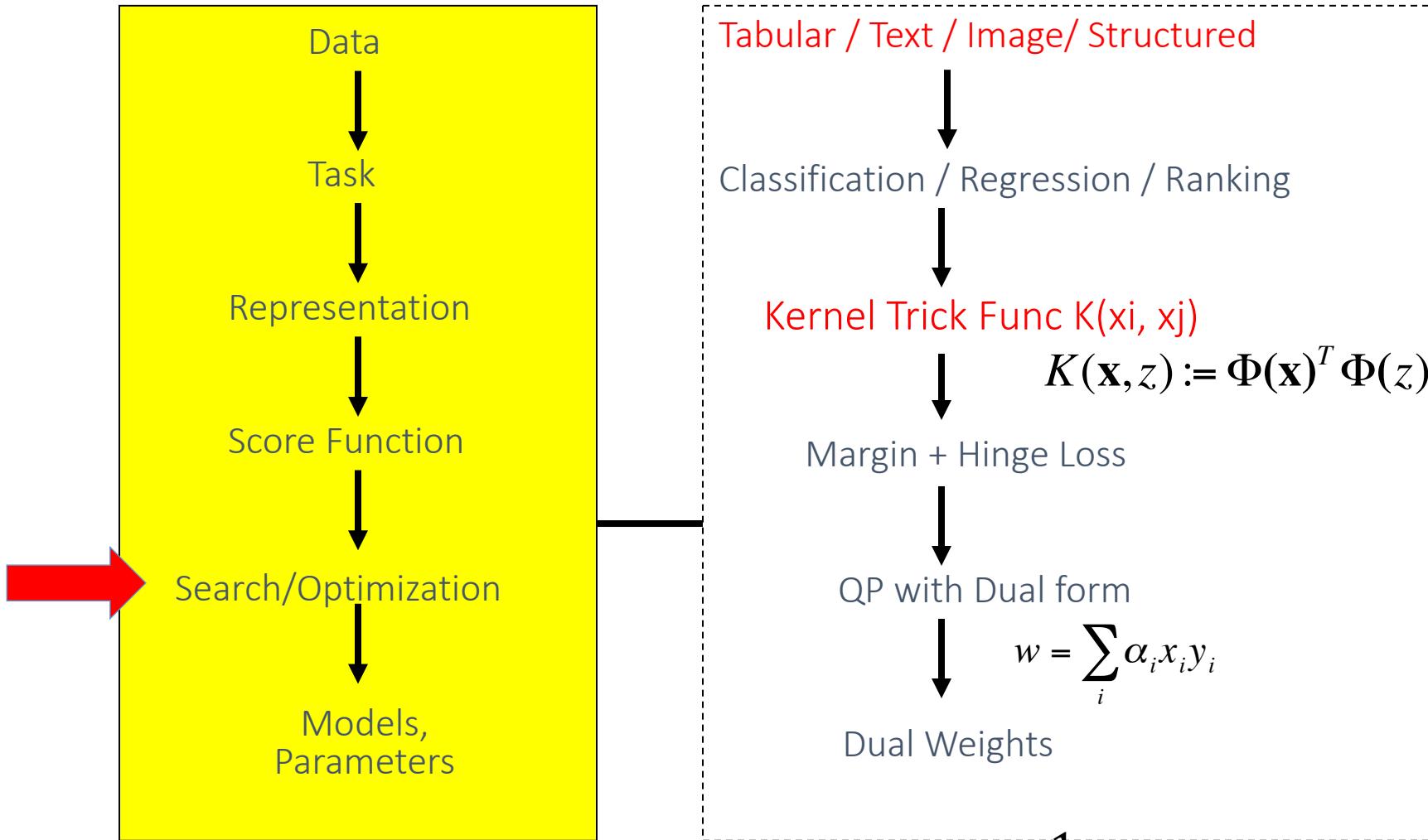
Department of Computer Science

Today Extra

❑ Optimization of SVM

- 
- ✓ SVM as QP
 - ✓ A simple example of constrained optimization
 - ✓ SVM Optimization with dual form
 - ✓ KKT condition
 - ✓ SMO algorithm for fast SVM dual optimization

This: Kernel Support Vector Machine



$$\begin{aligned} & \underset{\mathbf{w}, b}{\operatorname{argmin}} \sum_{i=1}^p w_i^2 + C \sum_{i=1}^n \varepsilon_i \\ \text{subject to } & \forall \mathbf{x}_i \in D_{\text{train}} : y_i (\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 - \varepsilon_i \end{aligned}$$

↗

$$\begin{aligned} & \max_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ & \sum_i \alpha_i y_i = 0, \quad \alpha_i \geq 0 \quad \forall i \end{aligned}$$

Optimization with Quadratic programming (QP)

Quadratic programming solves optimization problems of the following form:

$$\min_U \frac{u^T R u}{2} + d^T u + c$$

subject to n inequality constraints:

$$\begin{aligned} a_{11}u_1 + a_{12}u_2 + \dots &\leq b_1 \\ \vdots &\quad \vdots \quad \vdots \\ a_{n1}u_1 + a_{n2}u_2 + \dots &\leq b_n \end{aligned}$$

and k equivalency constraints:

$$\begin{aligned} a_{n+1,1}u_1 + a_{n+1,2}u_2 + \dots &= b_{n+1} \\ \vdots &\quad \vdots \quad \vdots \\ a_{n+k,1}u_1 + a_{n+k,2}u_2 + \dots &= b_{n+k} \end{aligned}$$

u → variable

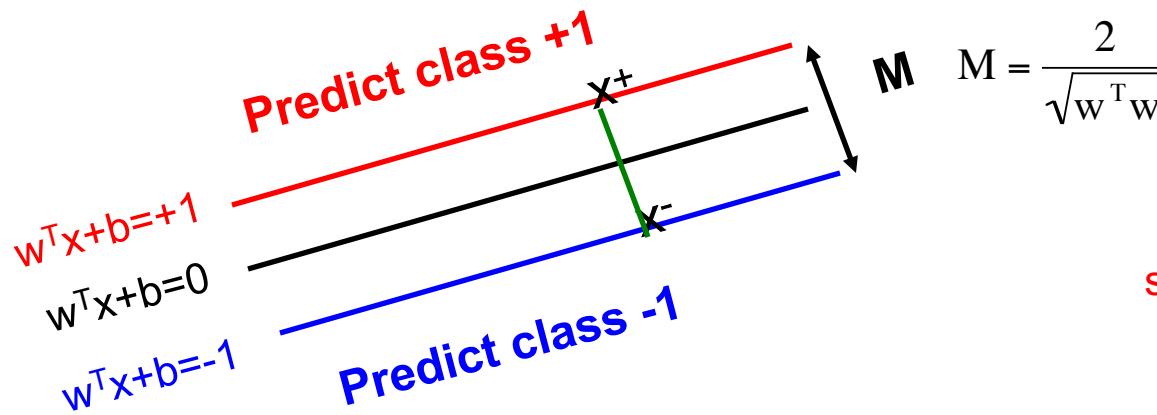
$f(u)$ → object

Quadratic term

$g_i(u)$ → Constraints

When a problem can be specified as a QP problem we can use solvers that are better than gradient descent or simulated annealing

SVM as a QP problem



$$\text{Min } (w^T w)/2$$

subject to the following inequality constraints:

For all x in class +1

$$w^T x + b \geq 1$$

For all x in class -1

$$w^T x + b \leq -1$$

}

A total of n constraints if we have n input samples

R as I matrix, d as zero vector, c as 0 value

$$\min_U \frac{u^T R u}{2} + d^T u + c$$

subject to n inequality constraints:

$$a_{11}u_1 + a_{12}u_2 + \dots \leq b_1$$

$$\vdots \quad \vdots \quad \vdots$$

$$a_{n1}u_1 + a_{n2}u_2 + \dots \leq b_n$$

and k equivalency constraints:

$$a_{n+1,1}u_1 + a_{n+1,2}u_2 + \dots = b_{n+1}$$

$$\vdots \quad \vdots \quad \vdots$$

$$a_{n+k,1}u_1 + a_{n+k,2}u_2 + \dots = b_{n+k}$$

Optimization Review: Ingredients

- Objective function
- Variables
- Constraints

**Find values of the variables
that minimize or maximize the objective function
while satisfying the constraints**

Today Extra

□ Optimization of SVM

- ✓ SVM as QP
- ✓ A simple example of constrained optimization and dual
- ✓ Optimization with dual form
- ✓ KKT condition
- ✓ SMO algorithm for fast SVM dual optimization

Optimization Review: Lagrangian Duality

- The Primal Problem

Primal:

$$\begin{aligned} \min_w \quad & f_0(w) \\ \text{s.t.} \quad & f_i(w) \leq 0, \quad i = 1, \dots, k \end{aligned}$$

The generalized Lagrangian:

“Method of Lagrange multipliers”
convert to a higher-dimensional problem

$$\mathcal{L}(w, \alpha) = f_0(w) + \sum_{i=1}^k \alpha_i f_i(w)$$

the α 's ($\alpha_i \geq 0$) are called the Lagrangian multipliers

Lemma:

$$\max_{\alpha, \alpha_i \geq 0} \mathcal{L}(w, \alpha) = \begin{cases} f_0(w) & \text{if } w \text{ satisfies primal constraints} \\ \infty & \text{o/w} \end{cases}$$

A re-written Primal:

$$\min_w \max_{\alpha, \alpha_i \geq 0} \mathcal{L}(w, \alpha)$$

Optimization Review: Lagrangian Duality, cont.

- Recall the Primal Problem:

$$\min_w \max_{\alpha, \alpha_i \geq 0} \mathcal{L}(w, \alpha)$$

- The Dual Problem:

$$\max_{\alpha, \alpha_i \geq 0} \min_w \mathcal{L}(w, \alpha)$$

- **Theorem (weak duality):**

$$d^* = \max_{\alpha, \alpha_i \geq 0} \min_w \mathcal{L}(w, \alpha) \leq \min_w \max_{\alpha, \alpha_i \geq 0} \mathcal{L}(w, \alpha) = p^*$$

- **Theorem (strong duality):**

Iff there exist a saddle point of $\mathcal{L}(w, \alpha)$

we have

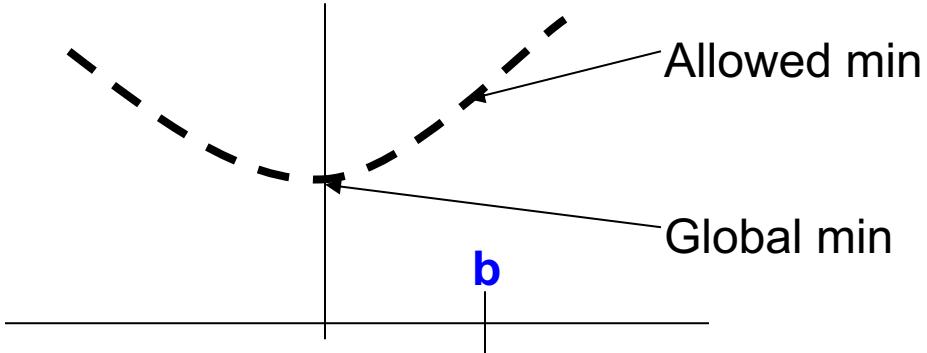
$$d^* = p^*$$

$\min_u u^2$ $s.t. u \geq b$

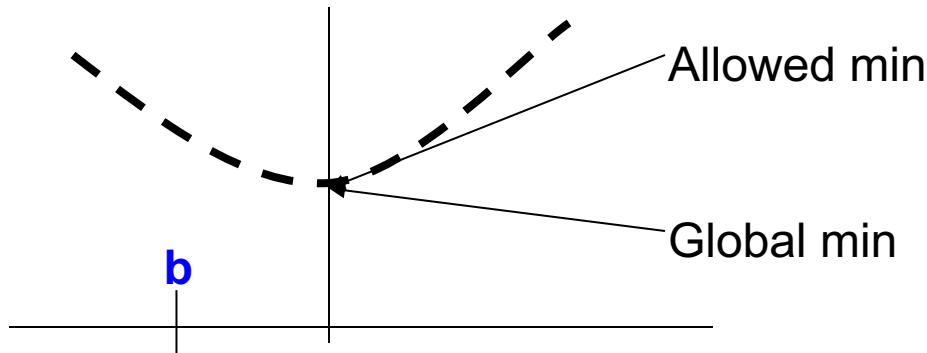
Optimization Review: Constrained Optimization

$$\begin{aligned} & \min_u u^2 \\ & \text{s.t. } u \geq b \end{aligned}$$

Case 1:



Case 2:

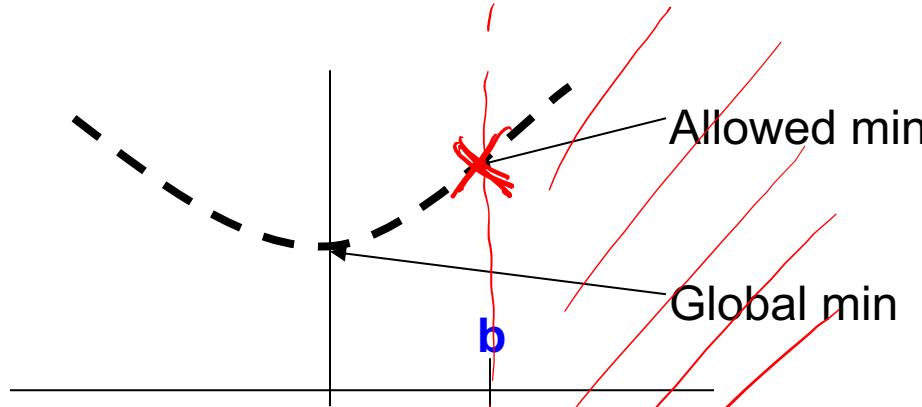


Optimization Review: Constrained Optimization

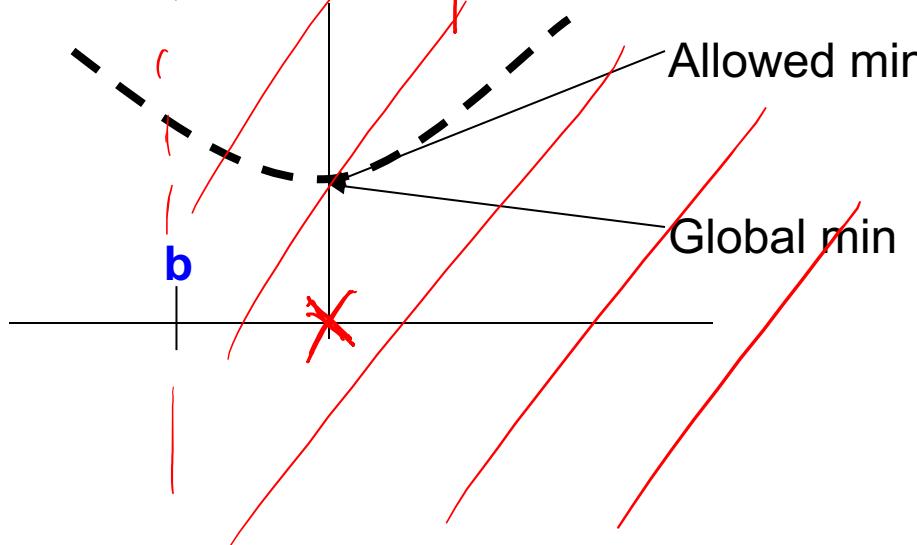
$$f(u) = u^2$$

$$\begin{aligned} & \min_u u^2 \\ & \text{s.t. } u \geq b \end{aligned}$$

Case 1:



Case 2:

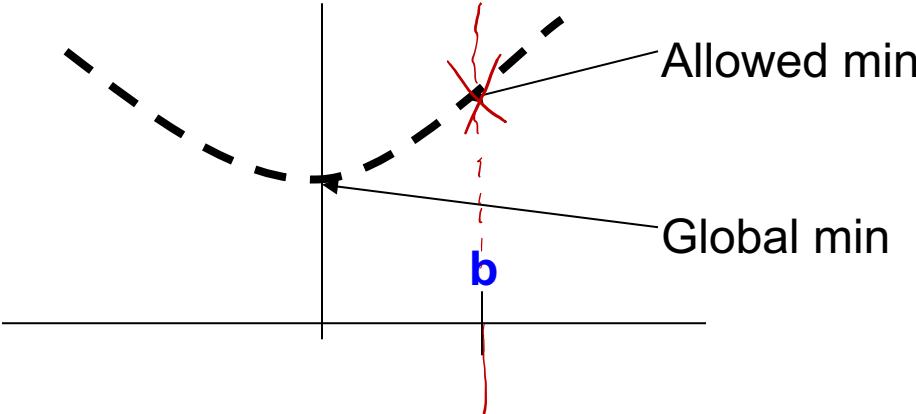


Optimization Review: Constrained Optimization

$$\begin{array}{|c|} \hline f(u) \\ \min_u u^2 \\ \text{s.t. } u \geq b \\ \hline \end{array}$$

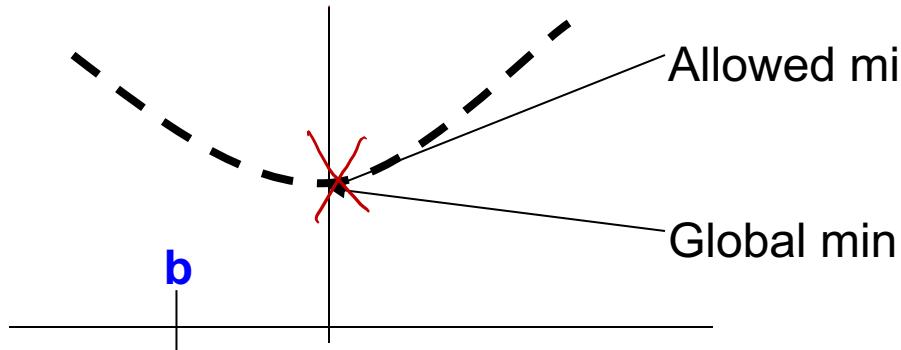
Subject to

Case 1:



$$b > 0$$
$$f(u) = b^2$$

Case 2:



$$b < 0$$
$$f(u) = 0$$

$$\begin{array}{l} \min_u u^2 \\ \text{s.t. } u \geq b \end{array}$$

$$D \left\{ \begin{array}{l} \min_u f_0(u) = u^2 \\ \text{s.t. } b - u \leq 0 \end{array} \right.$$

Primal problem

$$\begin{array}{l} \min_u u^2 \\ \text{s.t. } u \geq b \end{array}$$

$$0 \left\{ \begin{array}{l} \min_u f_0(u) = u^2 \\ \text{s.t. } b - u \leq 0 \end{array} \right.$$

\rightarrow multiplier variable

$$\textcircled{2} \quad L(u, \lambda) = u^2 + \underbrace{\lambda(b-u)}_{\geq 0 \leq 0}$$

$$\begin{array}{l} \min_u u^2 \\ \text{s.t. } u \geq b \end{array}$$

$$0 \left\{ \begin{array}{l} \min_u f_0(u) = u^2 \\ \text{s.t. } b - u \leq 0 \end{array} \right.$$

$$\textcircled{2} \quad L(u, \alpha) = u^2 + \underbrace{\alpha(b-u)}_{\geq 0 \leq 0}$$

$$\textcircled{3} \quad \frac{\partial L(u, \alpha)}{\partial u} = 2u - \alpha = 0$$

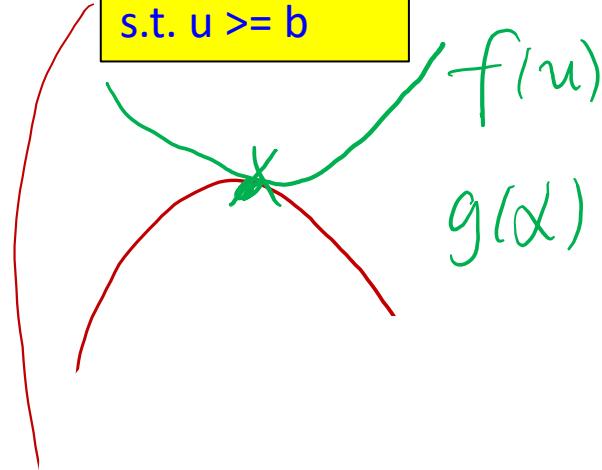
$$u = \frac{\alpha}{2}$$

$$\rightarrow \arg \min_u L(u, \alpha)$$

$$\begin{aligned} & \min_u u^2 \\ & \text{s.t. } u \geq b \end{aligned}$$

g(\alpha) = L(u, \alpha) = \frac{\alpha^2}{4} + \alpha(b - \frac{\alpha}{2})

$$\begin{array}{l} \min_u u^2 \\ \text{s.t. } u \geq b \end{array}$$

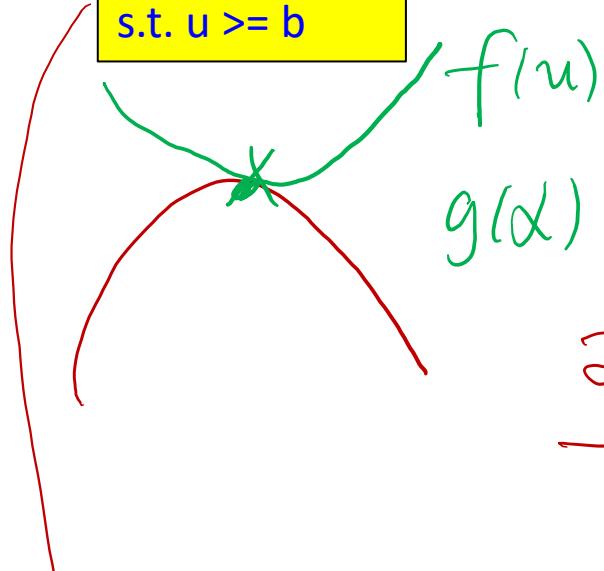


$$g(x) = L(u, x) = \frac{x^2}{4} + x(b - \frac{x}{2})$$

$$u = \frac{x}{2}$$

$$g(x) = -\frac{x^2}{4} + bx$$

$$\begin{array}{l} \min_u u^2 \\ \text{s.t. } u \geq b \end{array}$$



$$g(\alpha) = L(u, \alpha) = \frac{\alpha^2}{4} + \alpha(b - \frac{\alpha}{2})$$

$$u = \frac{\alpha}{2}$$

$$g(\alpha) = -\frac{\alpha^2}{4} + b\alpha$$

$$\frac{\partial g(\alpha)}{\partial \alpha} = \underline{-\frac{\alpha}{2} + b} = 0, \alpha \geq 0$$

$$\begin{aligned} & \min_u u^2 \\ \text{s.t. } & u \geq b \end{aligned}$$

$$g(\alpha) = L(u, \alpha) = \frac{\alpha^2}{4} + \alpha(b - \frac{\alpha}{2})$$

$f(u)$

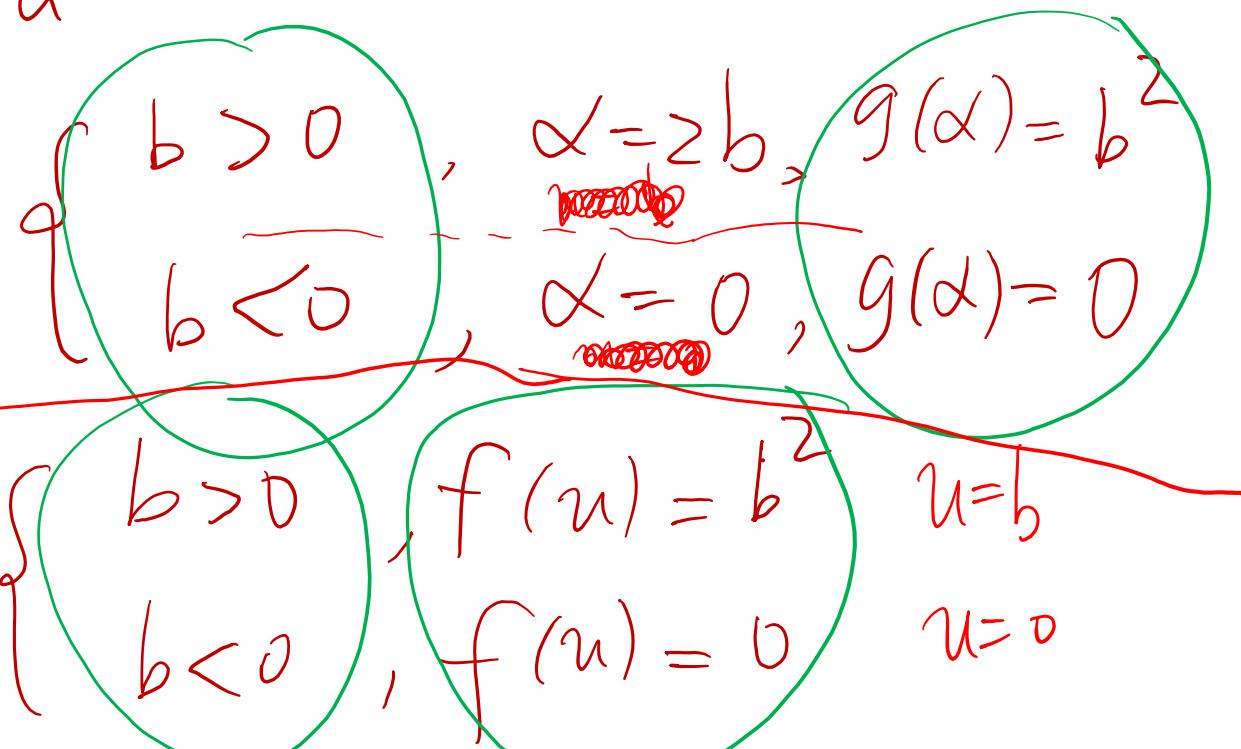
$$u = \frac{\alpha}{2}$$

$$g(\alpha) = -\frac{\alpha^2}{4} + b\alpha$$

$$\frac{\partial g(\alpha)}{\partial \alpha} = -\frac{\alpha}{2} + b = 0, \alpha \geq 0$$

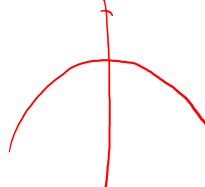
\Rightarrow
Dual

\Rightarrow
Primal



$$\text{Primal} : \min_w \max_{\alpha} L(w, \alpha)$$

$$\begin{aligned} \text{Dual} &: \max_{\alpha} \min_w L(w, \alpha) \\ &\Rightarrow \max_{\alpha} g(\alpha) \end{aligned}$$

$$f(u) : \begin{cases} \min u^2 \\ \text{s.t. } u \geq b \end{cases}$$


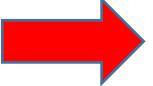
$$g(\alpha) : \begin{cases} \max -\frac{\alpha^2}{4} + b\alpha = \max \left\{ -\underbrace{(\alpha - b)^2}_{\geq 0} + b^2 \right\} \\ \text{s.t. } \alpha \geq 0 \end{cases}$$

$$\begin{cases} \text{if } b > 0, \quad u = b, \quad g = b^2 \\ \text{if } b \leq 0, \quad \alpha = 0, \quad g = 0 \end{cases}$$

$$\Rightarrow \alpha(b-u) = 0 \quad \text{KKT condition}$$

Today Extra

□ Optimization of SVM

- ✓ SVM as QP
 - ✓ A simple example of constrained optimization
 - ✓ SVM Optimization with dual form
 - ✓ KKT condition
 - ✓ SMO algorithm for fast SVM dual optimization
- 

$$\min_{w,b} \max_{\alpha} \frac{w^T w}{2} - \sum_i \alpha_i [(w^T x_i + b) y_i - 1]$$

$$\alpha_i \geq 0 \quad \forall i$$

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w - \sum_i^{\text{train}} \alpha_i x_i y_i = 0$$

$$\min_{w,b} \max_{\alpha} \left\{ \frac{w^T w}{2} - \sum_i \alpha_i [(w^T x_i + b) y_i - 1] \right\} \Rightarrow \max_{\alpha} \min_{w,b} L(w, b, \alpha)$$

$\alpha_i \geq 0$ $\forall i$

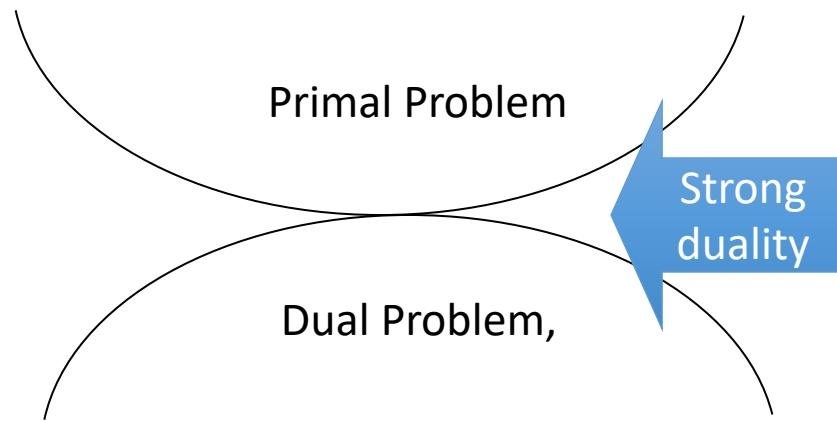
$$\begin{cases} \frac{\partial L}{\partial w} = 0 \Rightarrow w - \sum_i \alpha_i x_i y_i = 0 \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum \alpha_i y_i = 0 \end{cases}$$

$$L_{primal} = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i (y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1)$$

$$\begin{aligned} L_{dual} &= \frac{1}{2} \left(\sum_i \alpha_i y_i \right)^T \left(\sum_j \alpha_j x_j y_j \right) - \sum_i \alpha_i y_i \left(\sum_j \alpha_j x_j y_j \right)^T x_i \\ &\quad - \underbrace{\sum_i \alpha_i y_i}_b + \underbrace{\sum_i \alpha_i}_c \\ &= \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j (x_i^T x_j) \end{aligned}$$

Optimization Review: Dual Problem

- Solving dual problem if the dual form is easier than primal form
- Need to change primal **minimization** to dual **maximization** (OR → Need to change primal **maximization** to dual **minimization**)
- Only valid when the original optimization problem is convex/concave (strong duality)



Today Extra

❑ Optimization of SVM

- ✓ SVM as QP
- ✓ A simple example of constrained optimization
- ✓ SVM Optimization with dual form
-  ✓ KKT condition
- ✓ SMO algorithm for fast SVM dual optimization

KKT Condition for Strong Duality

minimize $\begin{bmatrix} f_0(x) \\ f_i(x) \leq 0, \\ h_i(x) = 0, \end{bmatrix}$
subject to $i = 1, \dots, m$
 $i = 1, \dots, p$

$$\min_w \max_{\alpha} L(w, \alpha)$$

Primal Problem

Dual Problem,

$$\max_{\alpha} \min_w L(w, \alpha)$$

Strong duality

Lagrangian: $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \rightarrow \mathbf{R}$, with $\text{dom } L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$$

complementary slackness: $\lambda_i f_i(x) = 0, i = 1, \dots, m$

Key for SVM Dual

Optimization Review: Lagrangian (even more general standard form)

standard form problem (not necessarily convex)

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && \begin{cases} f_i(x) \leq 0, & i = 1, \dots, m \\ h_i(x) = 0, & i = 1, \dots, p \end{cases} \end{aligned}$$

variable $x \in \mathbf{R}^n$, domain \mathcal{D} , optimal value p^*

Lagrangian: $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \rightarrow \mathbf{R}$, with $\text{dom } L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$

Optimization Review: Lagrange dual function

Lagrange dual function: $g : \mathbf{R}^m \times \mathbf{R}^p \rightarrow \mathbf{R}$,

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\ &= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right) \end{aligned}$$

g is concave, can be $-\infty$ for some λ, ν

lower bound property: if $\lambda \succeq 0$, then $g(\lambda, \nu) \leq p^*$

proof: if \tilde{x} is feasible and $\lambda \succeq 0$, then

Inf(.): greatest
lower bound

$$f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu)$$

minimizing over all feasible \tilde{x} gives $p^* \geq g(\lambda, \nu)$

Optimization Review:

Complementary slackness

assume strong duality holds, x^* is primal optimal, (λ^*, ν^*) is dual optimal

$\inf(\cdot)$: greatest lower bound

$$\begin{aligned} f_0(x^*) = g(\lambda^*, \nu^*) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \right) \\ &\leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

hence, the two inequalities hold with equality

- x^* minimizes $L(x, \lambda^*, \nu^*)$
- $\lambda_i^* f_i(x^*) = 0$ for $i = 1, \dots, m$ (known as complementary slackness):

$$\lambda_i^* > 0 \implies f_i(x^*) = 0, \quad f_i(x^*) < 0 \implies \lambda_i^* = 0$$

Optimization Review:

Complementary slackness

assume strong duality holds, x^* is primal optimal, (λ^*, ν^*) is dual optimal

$$\begin{aligned}
 & f_0(x^*) = g(\lambda^*, \nu^*) \\
 \text{obj} \Rightarrow & f_0(x^*) \\
 & \left. \begin{array}{l} u^* \geq b \\ b < 0 \end{array} \right\} u^* = b \\
 & \left. \begin{array}{l} u^* = 0 \\ u^* = 0 \end{array} \right\} u^* = 0 \\
 & \left. \begin{array}{l} \lambda^* \geq 0 \\ \lambda^* = 0 \end{array} \right\} \lambda^* = 0 \\
 & \left. \begin{array}{l} \nu_i^* \geq 0 \\ \nu_i^* = 0 \end{array} \right\} \nu_i^* = 0
 \end{aligned}
 \quad \begin{aligned}
 & \inf (\cdot) : \text{greatest lower bound} \\
 & = \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \right) \\
 & \leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) \\
 & \quad \left. \begin{array}{c} \geq 0 \\ \leq 0 \end{array} \right\} = 0
 \end{aligned}$$

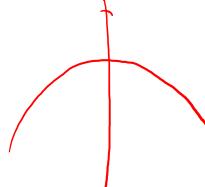
hence, the two inequalities hold with equality

- x^* minimizes $L(x, \lambda^*, \nu^*)$
- $\lambda_i^* f_i(x^*) = 0$ for $i = 1, \dots, m$ (known as complementary slackness):

$$\textcircled{1} \quad \lambda_i^* = 0$$

$$\textcircled{2} \quad \lambda_i^* > 0$$

$$\lambda_i^* > 0 \implies f_i(x^*) = 0, \quad f_i(x^*) < 0 \implies \lambda_i^* = 0$$

$$f(u) : \begin{cases} \min u^2 \\ \text{s.t. } u \geq b \end{cases}$$


$$g(\alpha) : \begin{cases} \max -\frac{\alpha^2}{4} + b\alpha = \max \left\{ -\underbrace{(\alpha - b)^2}_{\geq 0} + b^2 \right\} \\ \text{s.t. } \alpha \geq 0 \end{cases}$$

$$\begin{cases} \text{if } b > 0, \quad u = b, \quad g = b^2 \\ \text{if } b \leq 0, \quad \alpha = 0, \quad g = 0 \end{cases}$$

$$\Rightarrow \alpha(b-u) = 0 \quad \text{KKT condition}$$

Optimization Review: Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i):

1. primal constraints: $f_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, p$
2. dual constraints: $\lambda \succeq 0$
3. complementary slackness: $\lambda_i f_i(x) = 0, i = 1, \dots, m$
4. gradient of Lagrangian with respect to x vanishes:



Key for SVM Dual

$$\nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + \sum_{i=1}^p \nu_i \nabla h_i(x) = 0$$

~~from page 5/17~~: if strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions

Dual formulation for linearly non separable case (soft SVM)

Substituting (1), (2), and (3) into the Lagrange, we have:

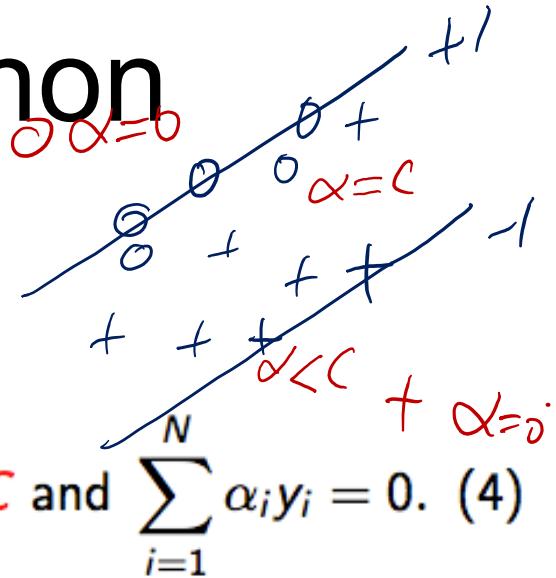
$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^N \alpha_i \alpha_k y_i y_k x_i^T x_k, \text{ with } 0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^N \alpha_i y_i = 0. \quad (4)$$

- $\hat{\alpha}_i > 0$: which implies $y_i(x_i^T \hat{\mathbf{w}} + \hat{b}) - 1 + \hat{\xi}_i = 0$ according to (5). These points are the *support vectors*.
 - $\hat{\xi}_i = 0$: which implies $\hat{\mu}_i > 0$ from (6) and so $\hat{\alpha}_i < C$ from (3). There are the support points which lie on the edge of the margin.
 - $\hat{\xi}_i > 0$: which implies $\hat{\mu}_i = 0$ from (6) and so $\hat{\alpha}_i = C$ from (3). There are the support points which violate the margin.
- $\hat{\alpha}_i = 0$: These points are not support vectors, which play no role in determining the hyperplane.

Dual formulation for linearly non separable case

Substituting (1), (2), and (3) into the Lagrange, we have:

$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^N \alpha_i \alpha_k y_i y_k x_i^T x_k, \text{ with } 0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^N \alpha_i y_i = 0. \quad (4)$$



- $\hat{\alpha}_i > 0$: which implies $y_i(x_i^T \hat{w} + \hat{b}) - 1 + \hat{\xi}_i = 0$ according to (5). These points are the *support vectors*.
 - $\hat{\xi}_i = 0$: which implies $\hat{\mu}_i > 0$ from (6) and so $\hat{\alpha}_i < C$ from (3). There are the support points which lie on the edge of the margin.
 - $\hat{\xi}_i > 0$: which implies $\hat{\mu}_i = 0$ from (6) and so $\hat{\alpha}_i = C$ from (3). There are the support points which violate the margin.
- $\hat{\alpha}_i = 0$: These points are not support vectors, which play no role in determining the hyperplane.

Support Vectors for the Soft-case

- Support vectors are

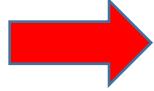
- Samples on the margin: $y_i(\mathbf{x}_i \cdot \mathbf{w} + b) = 1,$
 $0 < \alpha_i < C$

- Sample violate (mostly inside the margin area):

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) < 1,$$
$$\alpha_i = C$$

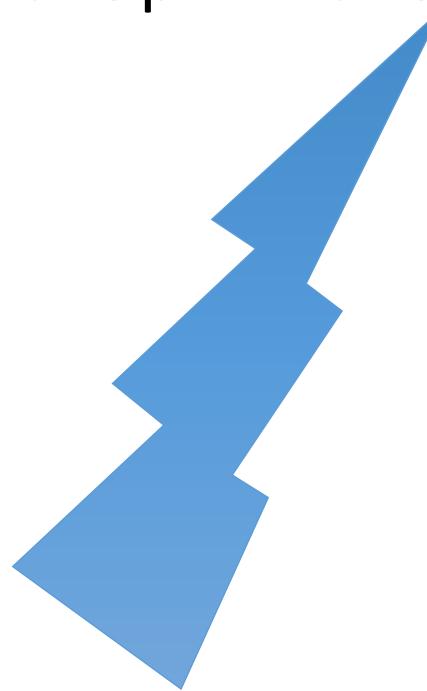
Today Extra

□ Optimization of SVM

- ✓ SVM as QP
 - ✓ A simple example of constrained optimization
 - ✓ SVM Optimization with dual form
 - ✓ KKT condition
 - ✓ SMO algorithm for fast SVM dual optimization
- 

Fast SVM Implementations

- SMO: Sequential Minimal Optimization
- SVM-Light
- LibSVM
- BSVM
-



J. Platt (1999),

Fast Training of Support Vector Machines Using Sequential Minimal Optimization

<https://pdfs.semanticscholar.org/d1fa/8485ad749d51e7470d801bc1931706597601.pdf>

SMO: Sequential Minimal Optimization

- Key idea
 - Divide the large QP problem of SVM into a series of smallest possible QP problems, which can be solved analytically and thus avoids using a time-consuming numerical QP in the loop (a kind of SQP method).
 - Space complexity: $O(n)$.
 - Since QP is greatly simplified, most time-consuming part of SMO is the evaluation of decision function, therefore it is very fast for linear SVM and sparse data.

SMO

- At each step, SMO chooses 2 Lagrange multipliers to jointly optimize, find the optimal values for these multipliers and updates the SVM to reflect the new optimal values.
- Three components
 - An analytic method to solve for the two Lagrange multipliers
 - A heuristic for choosing which (next) two multipliers to optimize
 - A method for computing b at each step, so that the KTT conditions are fulfilled for both the two examples (corresponding to the two multipliers)

Choosing Which Multipliers to Optimize

- First multiplier
 - Iterate over the entire training set, and find an example that violates the KTT condition.
- Second multiplier
 - Maximize the size of step taken during joint optimization.
 - $|E_1 - E_2|$, where E_i is the error on the i -th example.

References

- Big thanks to Prof. Ziv Bar-Joseph and Prof. Eric Xing @ CMU for allowing me to reuse some of his slides
- Elements of Statistical Learning, by Hastie, Tibshirani and Friedman
- Prof. Andrew Moore @ CMU's slides
- Tutorial slides from Dr. Tie-Yan Liu, MSR Asia
- A Practical Guide to Support Vector Classification Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, 2003-2010
- Tutorial slides from Stanford “Convex Optimization I — Boyd & Vandenberghe