

# Dual Vector Support Machine

这次介绍另外一种SVM，先回顾一下linear(plus transform) SVM：

## Non-Linear Support Vector Machine Revisited

$$\begin{aligned} \min_{b, \mathbf{w}} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} \\ \text{s. t.} \quad & y_n (\mathbf{w}^T \underbrace{\mathbf{z}_n}_{\Phi(\mathbf{x}_n)} + b) \geq 1, \\ & \text{for } n = 1, 2, \dots, N \end{aligned}$$

### Non-Linear Hard-Margin SVM

- 1  $Q = \begin{bmatrix} 0 & \mathbf{0}_{\tilde{d}}^T \\ \mathbf{0}_{\tilde{d}} & I_{\tilde{d}} \end{bmatrix}; \mathbf{p} = \mathbf{0}_{\tilde{d}+1};$   
 $\mathbf{a}_n^T = y_n [1 \quad \mathbf{z}_n^T]; c_n = 1$
- 2  $\begin{bmatrix} b \\ \mathbf{w} \end{bmatrix} \leftarrow \text{QP}(Q, \mathbf{p}, A, \mathbf{c})$
- 3 return  $b \in \mathbb{R} \ \& \ \mathbf{w} \in \mathbb{R}^{\tilde{d}}$  with  
 $g_{\text{SVM}}(\mathbf{x}) = \text{sign}(\mathbf{w}^T \Phi(\mathbf{x}) + b)$

- demanded: **not many** (large-margin), but **sophisticated** boundary (feature transform)
- QP with  $\tilde{d} + 1$  variables and  $N$  constraints  
 —challenging if  $\tilde{d}$  large, **or infinite?! :-)**

goal: SVM **without dependence on  $\tilde{d}$**

特点：凸优化quadratic problem

- $\tilde{d} + 1$  variables,  $\tilde{d}$  是由  $d$  进行 *linear transformation* 之后得到的维度
- $N$  constraints, QP约束条件

我们引入一种类似的SVM：

- $N$  variables
- $N+1$  constraints

方法：lagrange multipliers  $\alpha_n$  类比regularization中引入 $\lambda$

## Lagrange Function

with Lagrange multipliers  ~~$\alpha_n$~~ ,  $\alpha_n$ ,

$$\mathcal{L}(b, \mathbf{w}, \boldsymbol{\alpha}) = \underbrace{\frac{1}{2} \mathbf{w}^T \mathbf{w}}_{\text{objective}} + \sum_{n=1}^N \alpha_n \underbrace{(1 - y_n (\mathbf{w}^T \mathbf{z}_n + b))}_{\text{constraint}}$$

等价于下面的问题：

## Claim

$$\text{SVM} \equiv \min_{b, \mathbf{w}} \left( \max_{\text{all } \alpha_n \geq 0} \mathcal{L}(b, \mathbf{w}, \alpha) \right) = \min_{b, \mathbf{w}} \left( \infty \text{ if violate ; } \frac{1}{2} \mathbf{w}^T \mathbf{w} \text{ if feasible} \right)$$

- any 'violating'  $(b, \mathbf{w})$ :  $\max_{\text{all } \alpha_n \geq 0} \left( \square + \sum_n \alpha_n (\text{some positive}) \right) \rightarrow \infty$
- any 'feasible'  $(b, \mathbf{w})$ :  $\max_{\text{all } \alpha_n \geq 0} \left( \square + \sum_n \alpha_n (\text{all non-positive}) \right) = \square$

证明等价性两步：

- violating: 可以趋于正无穷
- feasible: 可以保持去拉格朗日项的结果

下面问题划归为如下：

$$\min_{b, w} \left( \max_{\text{all } \alpha_n \geq 0} L(b, w, \alpha) \right) \quad (1)$$

进行一步对于max的放缩: 我们固定一个任意的 $\alpha'$

$$\min_{b, w} \left( \max_{\text{all } \alpha_n \geq 0} L(b, w, \alpha) \right) \geq \min_{b, w} (L(b, w, \alpha')) \quad (2)$$

$$\min_{b, w} \left( \max_{\text{all } \alpha_n \geq 0} L(b, w, \alpha) \right) \geq \max_{\text{all } \alpha'_n \geq 0} \left( \min_{b, w} (L(b, w, \alpha')) \right) \quad (3)$$

可是 (3) 式还不够强, 我们希望有“=”这种强大的条件

**QP问题中, 如果满足如下条件, 那么等号成立**

- 凸优化问题
- 有解, 在Z空间里面是separable
- 有线性约束

——called constraint qualification

那么一定存在最佳解, 对于不等式左右两边都是成立的! 于是我们考虑max(min)问题

$$\max_{\text{all } \alpha_n \geq 0} \left( \min_{b, \mathbf{w}} \underbrace{\frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_{n=1}^N \alpha_n (1 - y_n (\mathbf{w}^T \mathbf{z}_n + b))}_{\mathcal{L}(b, \mathbf{w}, \alpha)} \right)$$

考虑无约束情况下的情况, 即这个问题的必要条件:  $\text{gradient}(b) = 0$

- inner problem 'unconstrained', at optimal:

$$\frac{\partial \mathcal{L}(b, \mathbf{w}, \alpha)}{\partial b} = 0 = - \sum_{n=1}^N \alpha_n y_n$$

- no loss of optimality if solving with constraint  $\sum_{n=1}^N \alpha_n y_n = 0$

$$0 = - \sum_{n=1}^N \alpha_n y_n \quad (4)$$

那么就有了新的约束条件, 简化问题:

but wait,  $b$  can be removed

$$\max_{\text{all } \alpha_n \geq 0, \sum y_n \alpha_n = 0} \left( \min_{b, \mathbf{w}} \frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_{n=1}^N \alpha_n (1 - y_n (\mathbf{w}^T \mathbf{z}_n)) - \sum_{n=1}^N \alpha_n y_n \cdot b \right)$$

再次  $\text{gradient}(\mathbf{w}) = 0$

我们有,

$$0 = \mathbf{w} - \sum_{n=1}^N \alpha_n y_n \mathbf{z}_n \quad (5)$$

but wait!

$$\begin{aligned} & \max_{\text{all } \alpha_n \geq 0, \sum y_n \alpha_n = 0, \mathbf{w} = \sum \alpha_n y_n \mathbf{z}_n} \left( \min_{b, \mathbf{w}} \frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_{n=1}^N \alpha_n - \mathbf{w}^T \mathbf{w} \right) \\ \iff & \max_{\text{all } \alpha_n \geq 0, \sum y_n \alpha_n = 0, \mathbf{w} = \sum \alpha_n y_n \mathbf{z}_n} -\frac{1}{2} \left\| \sum_{n=1}^N \alpha_n y_n \mathbf{z}_n \right\|^2 + \sum_{n=1}^N \alpha_n \end{aligned}$$

于是著名的KKT条件就出来了:

## KKT Optimality Conditions

$$\max_{\text{all } \alpha_n \geq 0, \sum y_n \alpha_n = 0, \mathbf{w} = \sum \alpha_n y_n \mathbf{z}_n} -\frac{1}{2} \left\| \sum_{n=1}^N \alpha_n y_n \mathbf{z}_n \right\|^2 + \sum_{n=1}^N \alpha_n$$

if primal-dual optimal  $(b, \mathbf{w}, \alpha)$ ,

- primal feasible:  $y_n (\mathbf{w}^T \mathbf{z}_n + b) \geq 1$
- dual feasible:  $\alpha_n \geq 0$
- dual-inner optimal:  $\sum y_n \alpha_n = 0$ ;  $\mathbf{w} = \sum \alpha_n y_n \mathbf{z}_n$
- primal-inner optimal (at optimal all 'Lagrange terms' disappear):

$$\alpha_n (1 - y_n (\mathbf{w}^T \mathbf{z}_n + b)) = 0$$

—called **Karush-Kuhn-Tucker (KKT) conditions**, necessary for optimality [& sufficient here]

will use **KKT** to 'solve'  $(b, \mathbf{w})$  from optimal  $\alpha$

$w$ 已经被干掉了, 现在就只剩下 $\alpha_n$ 了, 对于原问题优化可以得到complimentary slackness:

$$\alpha_n (1 - y_n (\mathbf{w}^T \mathbf{z}_n + b)) \quad (6)$$

KKT条件是一个充要条件\*\*

然后我们转化成标准的对偶支持向量机模型: 关于 $w$ 的条件先不考虑

standard hard-margin SVM dual

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \alpha_n \alpha_m y_n y_m \mathbf{z}_n^T \mathbf{z}_m - \sum_{n=1}^N \alpha_n \\ \text{subject to} \quad & \sum_{n=1}^N y_n \alpha_n = 0; \\ & \alpha_n \geq 0, \text{ for } n = 1, 2, \dots, N \end{aligned}$$

(convex) QP of  $N$  variables &  $N + 1$  constraints, as promised

QP问题可以解决:

## Dual SVM with QP Solver

optimal  $\alpha = ?$

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \alpha_n \alpha_m y_n y_m \mathbf{z}_n^T \mathbf{z}_m \\ & - \sum_{n=1}^N \alpha_n \\ \text{subject to} \quad & \sum_{n=1}^N y_n \alpha_n = 0; \\ & \alpha_n \geq 0, \\ & \text{for } n = 1, 2, \dots, N \end{aligned}$$

optimal  $\alpha \leftarrow \text{QP}(\mathbf{Q}, \mathbf{p}, \mathbf{A}, \mathbf{c})$

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \alpha^T \mathbf{Q} \alpha + \mathbf{p}^T \alpha \\ \text{subject to} \quad & \mathbf{a}_i^T \alpha \geq c_i, \\ & \text{for } i = 1, 2, \dots \end{aligned}$$

- $q_{n,m} = y_n y_m \mathbf{z}_n^T \mathbf{z}_m$
- $\mathbf{p} = -\mathbf{1}_N$
- $\mathbf{a}_{\geq} = \mathbf{y}, \mathbf{a}_{\leq} = -\mathbf{y};$   
 $\mathbf{a}_n^T = n\text{-th unit direction}$
- $c_{\geq} = 0, c_{\leq} = 0; c_n = 0$

不过很多解二次规划问题不用转化=为两个条件，具体情况看软件叭

这里要注意一个问题:

$\mathbf{Q}_D$  is too dense! (7)

所以我们需要特别的程序来解这个结果（看你用的lib叭233）

## KKT conditions

if primal-dual optimal  $(b, \mathbf{w}, \alpha)$ ,

- primal feasible:  $y_n(\mathbf{w}^T \mathbf{z}_n + b) \geq 1$
- dual feasible:  $\alpha_n \geq 0$
- dual-inner optimal:  $\sum y_n \alpha_n = 0$ ;  $\mathbf{w} = \sum \alpha_n y_n \mathbf{z}_n$
- primal-inner optimal (at optimal all 'Lagrange terms' disappear):

$$\alpha_n(1 - y_n(\mathbf{w}^T \mathbf{z}_n + b)) = 0 \text{ (complementary slackness)}$$

- optimal  $\alpha \implies$  optimal  $\mathbf{w}$ ? easy above!
- optimal  $\alpha \implies$  optimal  $b$ ? a range from primal feasible & equality from comp. slackness if one  $\alpha_n > 0 \Rightarrow b = y_n - \mathbf{w}^T \mathbf{z}_n$

comp. slackness:

$$\alpha_n > 0 \Rightarrow \text{on fat boundary (SV!)}$$

于是对于  $\alpha_n > 0$  情况都是 support vector

- only SV needed to compute  $\mathbf{w}$ :  $\mathbf{w} = \sum_{n=1}^N \alpha_n y_n \mathbf{z}_n = \sum_{\text{SV}} \alpha_n y_n \mathbf{z}_n$
- only SV needed to compute  $b$ :  $b = y_n - \mathbf{w}^T \mathbf{z}_n$  with any SV  $(\mathbf{z}_n, y_n)$

我们比较一下SVM和PLA，发现 $\mathbf{w}$ 都可以被data表示出来：

SVM中是通过support vector的线性组合表示出来，而PLA中是通过犯错误的点来表示出来：

### SVM

$$\mathbf{w}_{\text{SVM}} = \sum_{n=1}^N \alpha_n (y_n \mathbf{z}_n)$$

$\alpha_n$  from dual solution

### PLA

$$\mathbf{w}_{\text{PLA}} = \sum_{n=1}^N \beta_n (y_n \mathbf{z}_n)$$

$\beta_n$  by # mistake corrections

$\mathbf{w}$  = linear combination of  $y_n \mathbf{z}_n$

- also true for GD/SGD-based LogReg/LinReg when  $\mathbf{w}_0 = \mathbf{0}$
- call  $\mathbf{w}$  'represented' by data

对于两种SVM的总结：

## Summary: Two Forms of Hard-Margin SVM

### Primal Hard-Margin SVM

$$\begin{aligned} \min_{b, \mathbf{w}} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} \\ \text{sub. to} \quad & y_n (\mathbf{w}^T \mathbf{z}_n + b) \geq 1, \\ & \text{for } n = 1, 2, \dots, N \end{aligned}$$

- $\tilde{d} + 1$  variables,  
     $N$  constraints  
    —suitable when  $\tilde{d} + 1$  small
- physical meaning: locate  
    **specially-scaled**  $(b, \mathbf{w})$

### Dual Hard-Margin SVM

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \alpha^T Q_D \alpha - \mathbf{1}^T \alpha \\ \text{s.t.} \quad & \mathbf{y}^T \alpha = 0; \\ & \alpha_n \geq 0 \text{ for } n = 1, \dots, N \end{aligned}$$

- $N$  variables,  
     $N + 1$  simple constraints  
    —suitable when  $N$  small
- physical meaning: locate  
    **SVs**  $(\mathbf{z}_n, y_n)$  & their  $\alpha_n$

both eventually result in optimal  $(b, \mathbf{w})$  for fattest hyperplane

$$g_{\text{SVM}}(\mathbf{x}) = \text{sign}(\mathbf{w}^T \Phi(\mathbf{x}) + b)$$

这里我们还是没有解决一个重要的事情：

计算 $Q_d$ 的时候计算的复杂度也和 $d \sim$ 有关

解决方法 **kernel method!**