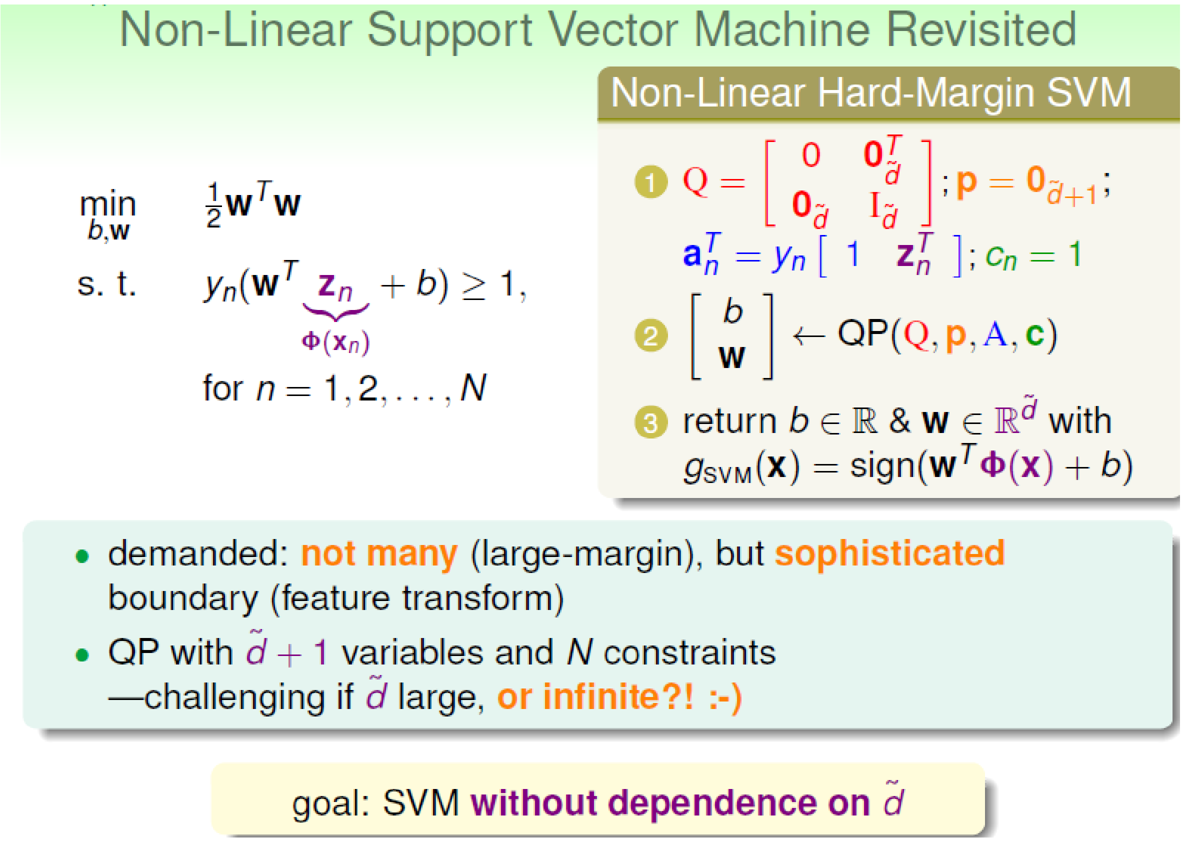
Dual Vector Support Machine

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



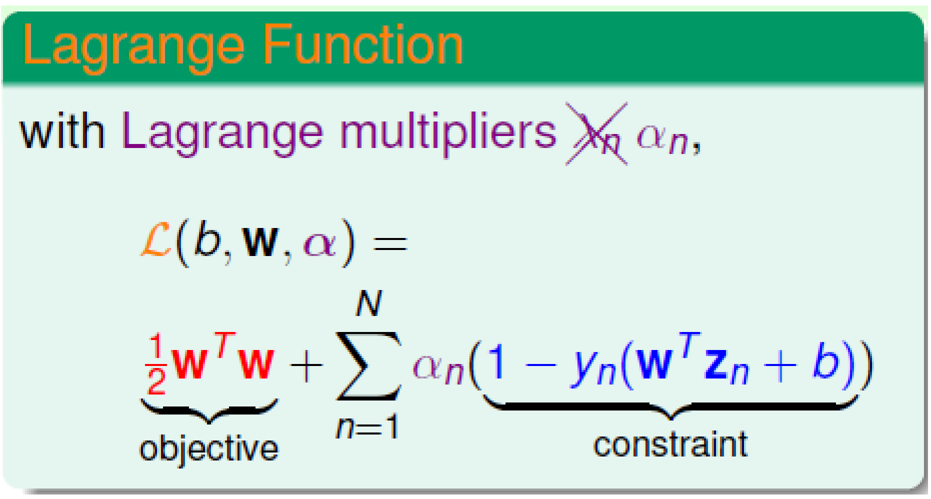
特点：凸优化quadratic problem

- $d^{\sim}+1$ variables , $d^{\sim}$是由d进行$lineartransformation$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 α_n 类比regularization中引入λ



等价于下面的问题：

Claim

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

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

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：

$$\min_{b,w}(\max_{all\ \alpha_n \geq 0} L(b, w, \alpha)) \tag{1}$$

进行一步对于max的放缩: 我们固定一个任意的α'

$$\min_{b,w}(\max_{all\ \alpha_n \geq 0} L(b, w, \alpha)) \geq \min_{b,w}(L(b, w, \alpha')) \tag{2}$$

$$\min_{b,w}(\max_{all\ \alpha_n \geq 0} L(b, w, \alpha)) \geq \max_{all\ \alpha_n' \geq 0} (\min_{b,w}(L(b, w, \alpha'))) \tag{3}$$

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 qualificaiton

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

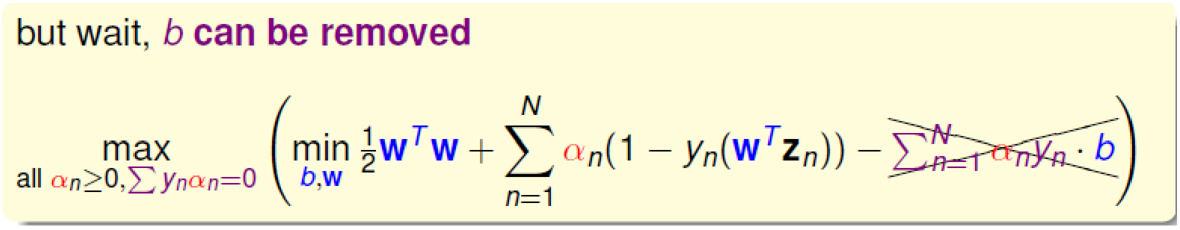
$$\max_{\text{all } \alpha_n \geq 0} \left(\underbrace{\min_{b,\mathbf{w}} \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n (1 - y_n(\mathbf{w}^T\mathbf{z}_n + b))}_{\mathcal{L}(b,\mathbf{w},\alpha)} \right)$$

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

- 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 $\sum_{n=1}^{N} \alpha_n y_n = 0$

$$0 = -\sum_{n=1}^{N} \alpha_n y_n \tag{4}$$

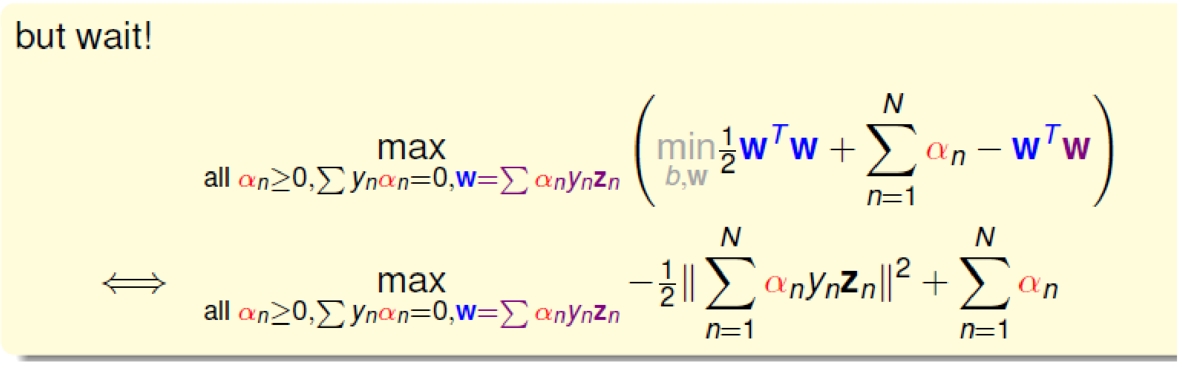
那么就有了新的约束条件，简化问题：



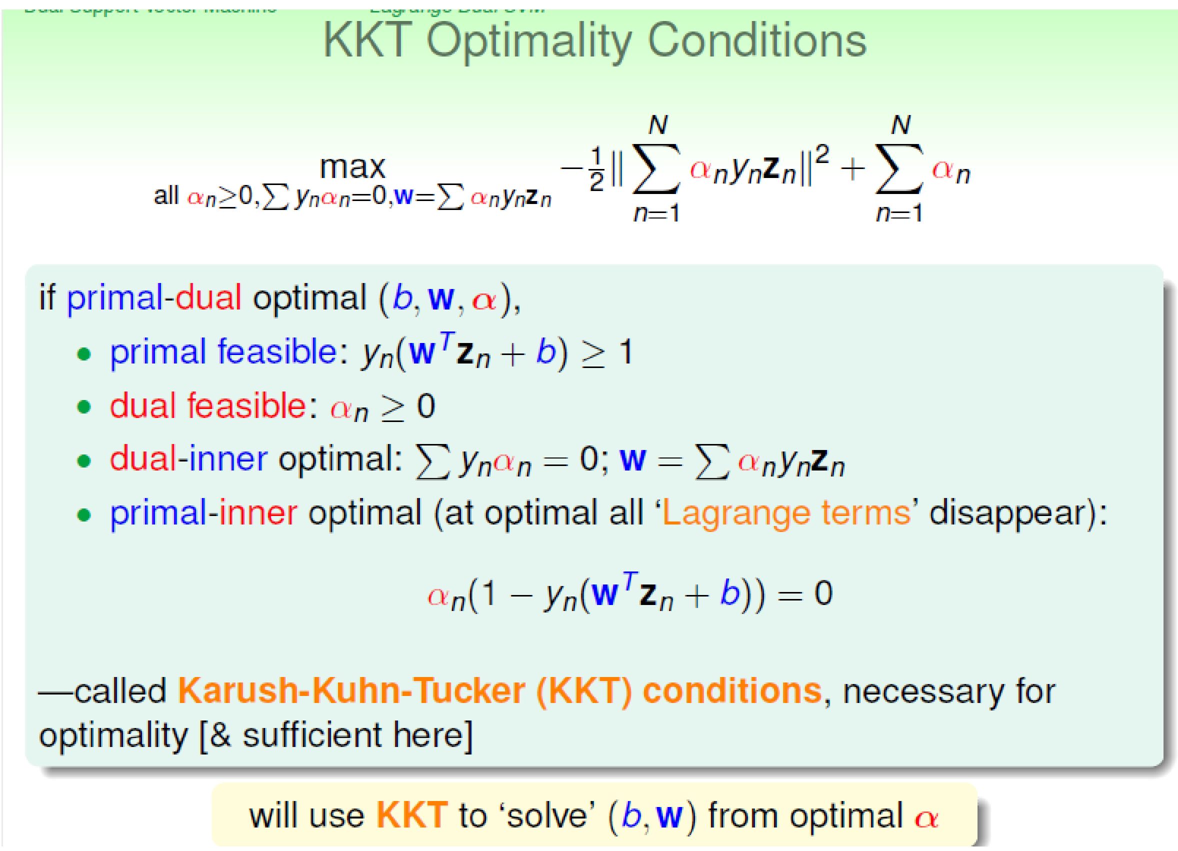
再次$gradient(w) = 0$

我们有，

$$0 = w - \sum_{n=1}^{N} \alpha_n y_n z_n \tag{5}$$



于是著名的KKT条件就出来了：



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

$$\alpha_n (1 - y_n (w^T z_n + b)) \tag{6}$$

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

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

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

QP问题可以解决：

Dual SVM with QP Solver

optimal $\alpha = ?$

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

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

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

- $q_{n,m} = y_n y_m \mathbf{z}_n^T \mathbf{z}_m$
- $\mathbf{p} = -\mathbf{1}_N$
- $\mathbf{a}_{\geq} = \mathbf{y}$, $\mathbf{a}_{\leq} = -\mathbf{y}$; $\mathbf{a}_n^T = n$-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：

$$
Q_D \ is\ too\ dense! \tag{7}
$$

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 $\alpha \Longrightarrow$ optimal $\mathbf{w}$? easy above!
- optimal $\alpha \Longrightarrow$ 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:
$\alpha_n > 0 \Rightarrow$ **on fat boundary (SV!)**

于是对于$\alpha_n > 0$情况的都是**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

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

α_n from **dual solution**

PLA

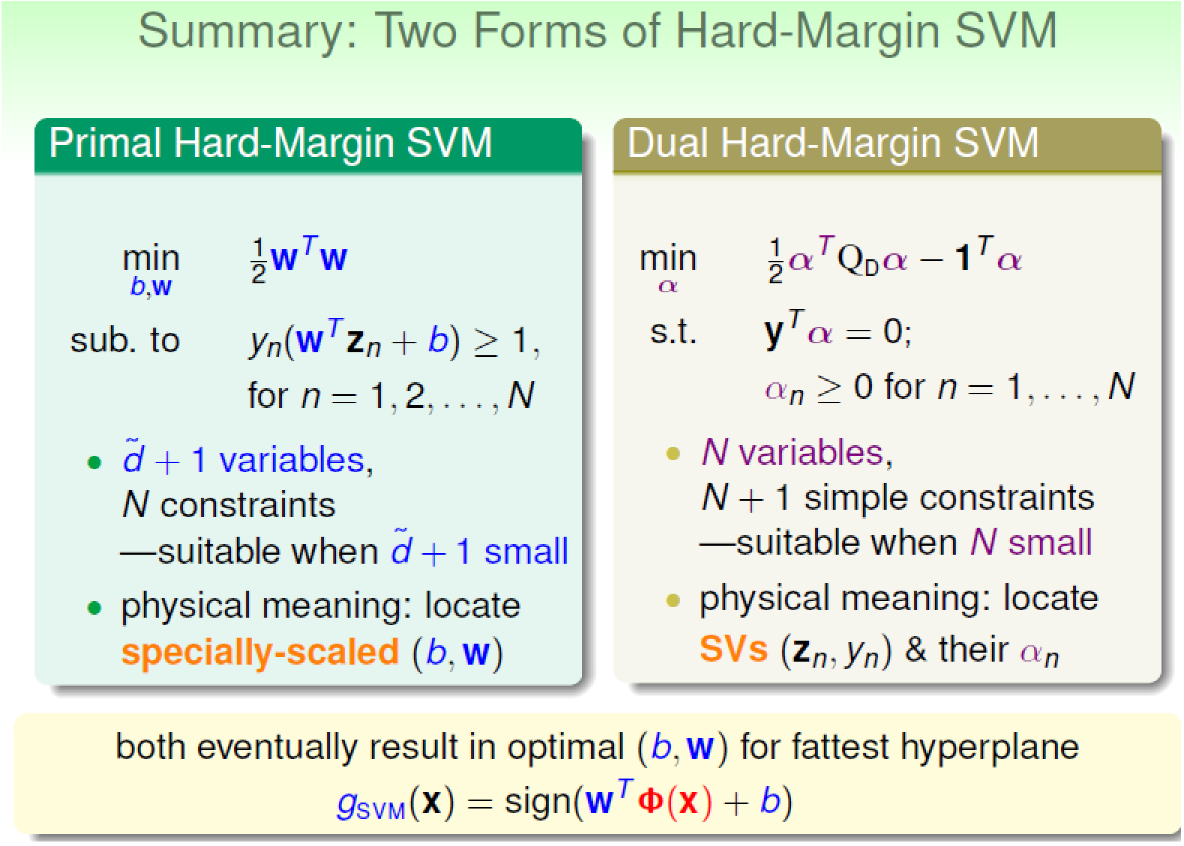
$$\mathbf{w}_{\text{PLA}} = \sum_{n=1}^{N} \beta_n(y_n\mathbf{z}_n)$$

β_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：



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$\tilde{d}$有关

解决方法 **kernel method**!