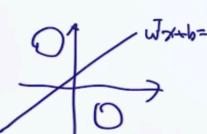


## 1. Linear Classifier

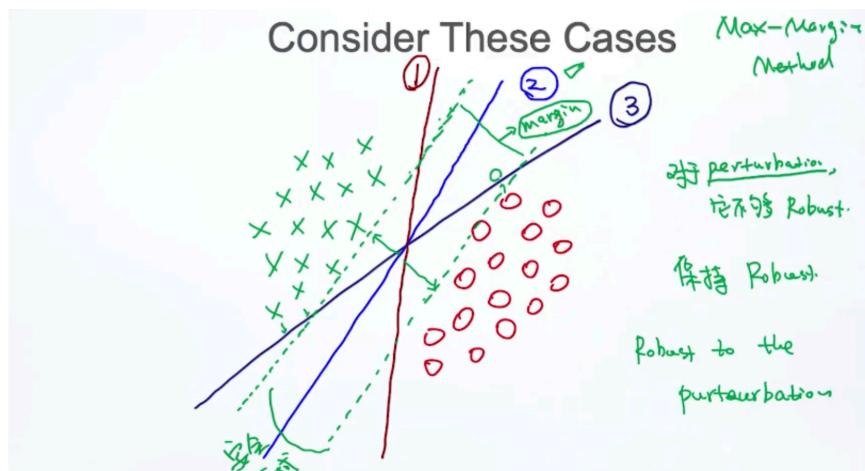
$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad y_i \in \{-1, 1\} \quad i=1, 2, \dots, n$$

$$\Theta = \{\omega, b\} \quad \underline{\omega^T x + b = 0} \leftarrow \text{Decision Boundary}$$

①  $\omega^T x_i + b \geq 0 \quad y_i = 1$   
 ②  $\omega^T x_i + b < 0 \quad y_i = -1$

$$\left. \begin{array}{l} \text{① } \omega^T x_i + b \geq 0 \\ \text{② } \omega^T x_i + b < 0 \end{array} \right\} \Rightarrow (\omega^T x_i + b) \cdot y_i \geq 0$$


□ PBR



## 2. SVM

### 2.1 Maximize the margin

Maximize the Margin

$$\begin{aligned} & \text{Margin} = |x_+ - x_-| \\ & \text{margin} = |\lambda w| \\ & = \lambda \|w\| \\ & = \frac{2}{w^T w} \|w\| \\ & = \boxed{\frac{2}{\|w\|}} \end{aligned}$$

$$\begin{aligned} & w^T x_+ + b = 1 \\ & w^T x_- + b = -1 \\ & x_+ = x_- + \lambda w \end{aligned} \quad \begin{aligned} & \text{Margin} = |x_+ - x_-| \\ & w^T (x_- + \lambda w) + b = 1 \\ & w^T x_- + \lambda w^T w + b = 1 \end{aligned}$$

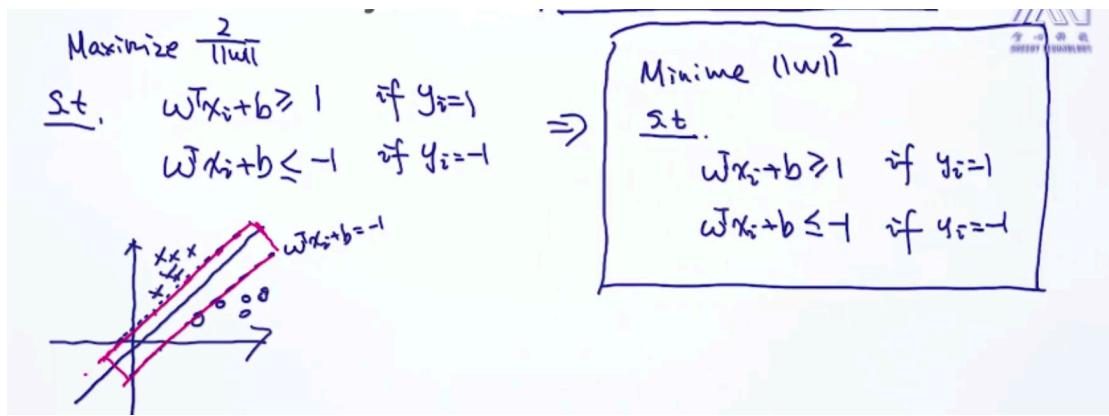
Maximize  $\boxed{\frac{2}{\|w\|}}$

S.t.

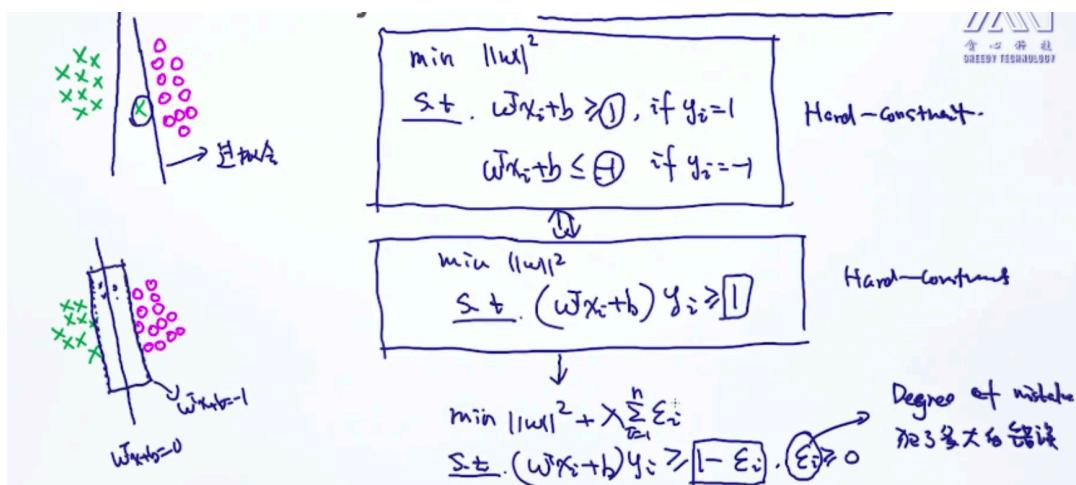
$$\lambda w^T w = 2$$

$$\lambda = \boxed{\frac{2}{\|w\|^2}}$$

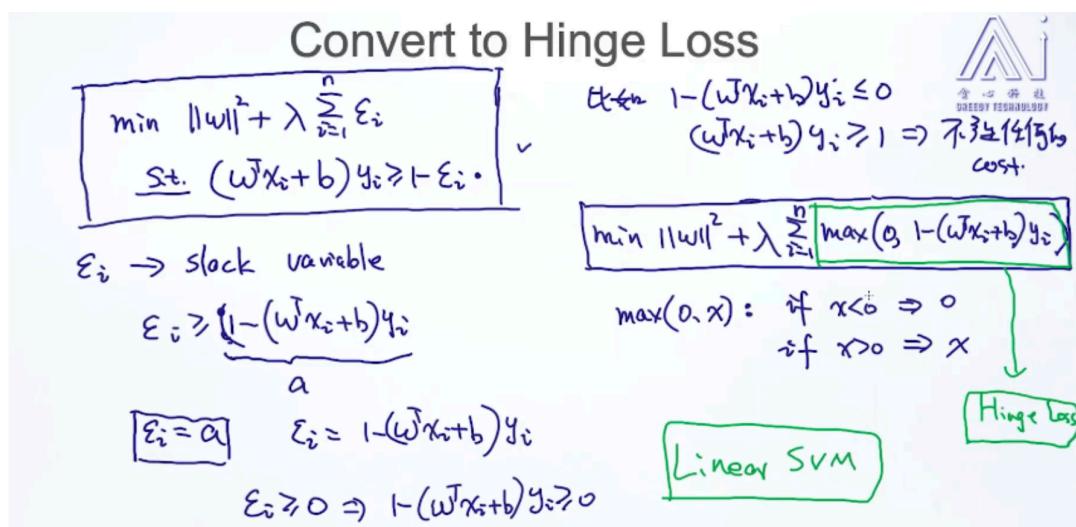
### 2.2 SVM 的目标函数: Hard constraint



## 2.3 SVM 的目标函数: Soft constraint



## 3. Hinge Loss



## 3.1 Gradient Descent for Hinge Loss Objective

$$\min \|w\|^2 + \lambda \sum_{i=1}^n \max(0, 1 - y_i(w^T x_i + b))$$

Regulation      Loss

$$\max(0, x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases}$$

$w^*, b^*$  (参数)

$i = 1 \dots n$

if  $1 - y_i(w^T x_i + b) \leq 0$

$w^* = w - \eta_+ \cdot 2w$

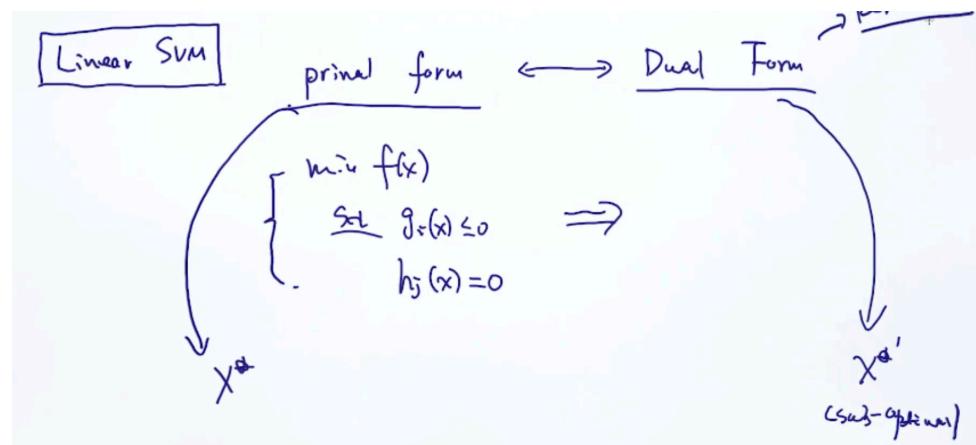
else  $w^* = w - \eta_+ (2w + \lambda \frac{\partial (1 - y_i(w^T x_i + b))}{\partial w})$

$b^* = b - \eta_+ (\lambda \frac{\partial (1 - y_i(w^T x_i + b))}{\partial b})$

SGD  
过程

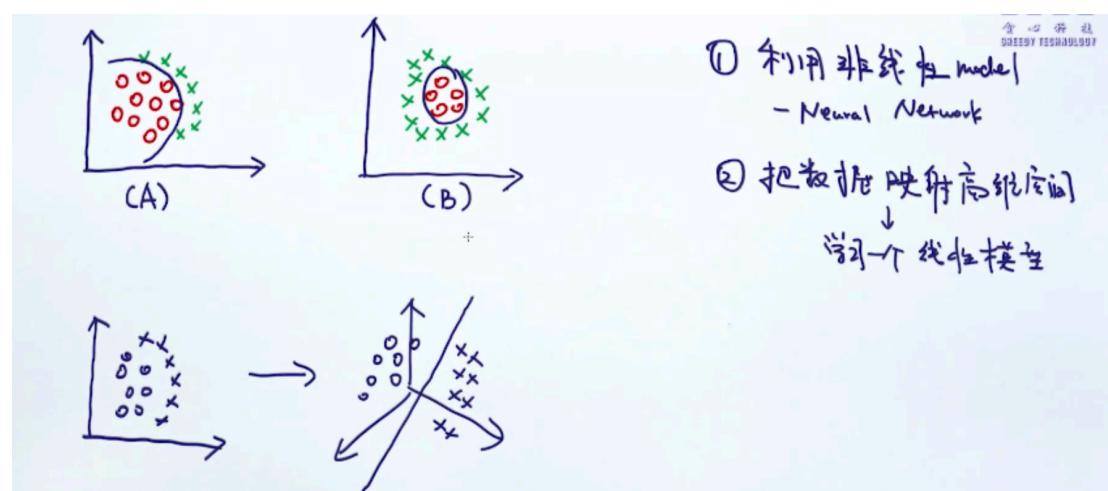
## 4. Linear SVM

### 4.1 Primal Form

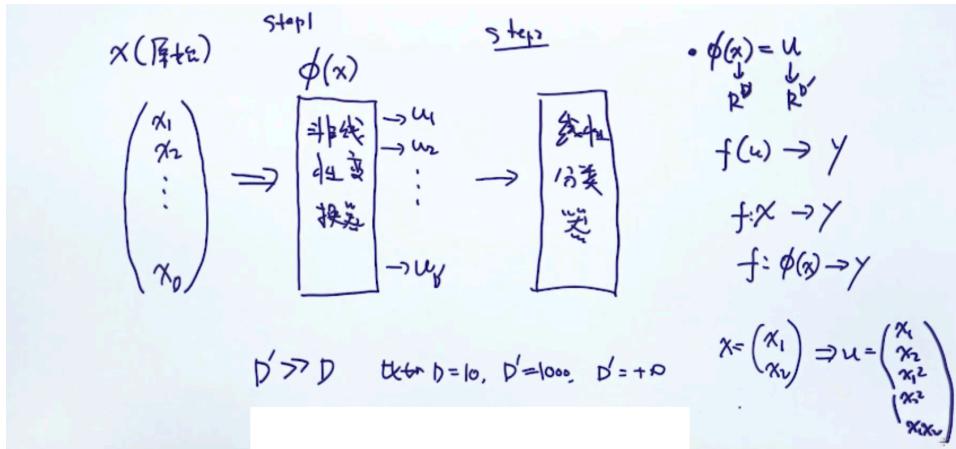


Kernel

### 4.2 linear SVM 的缺点

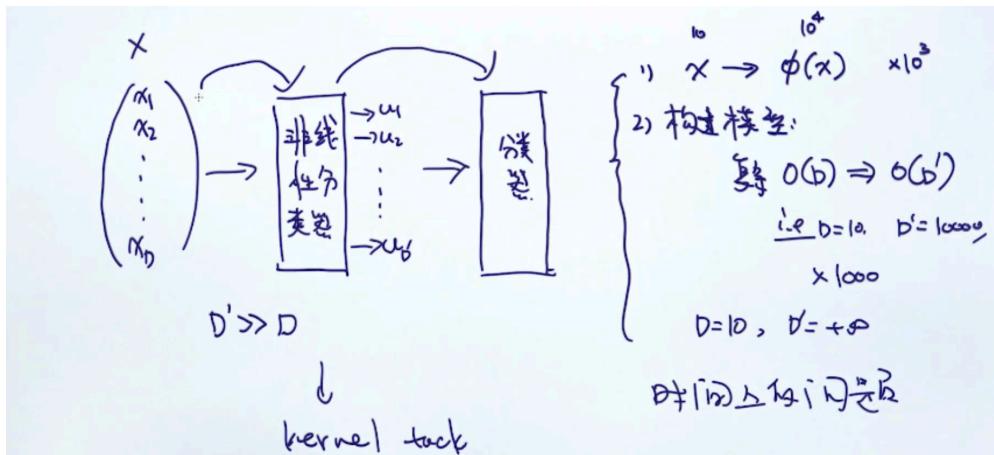


## 5. Mapping Feature to High Dimensional Space



缺点：时间复杂度提升了

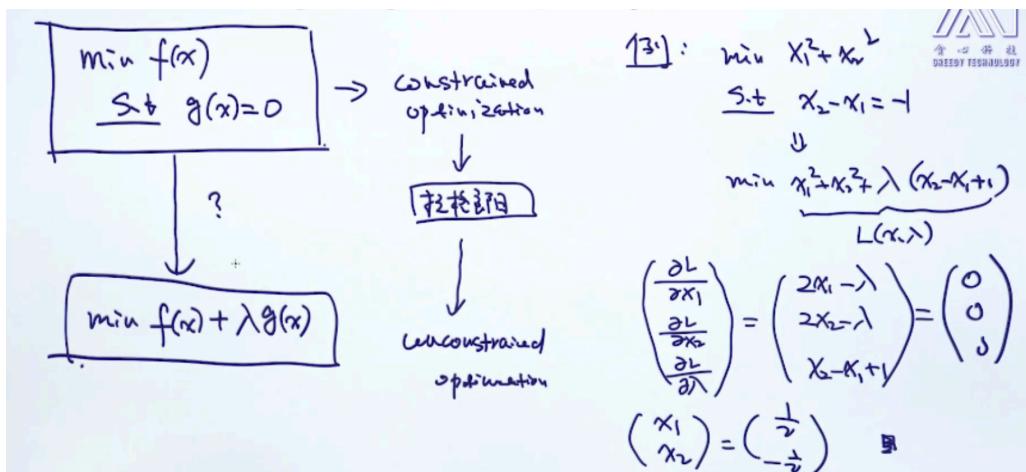
- 1) 把原始特征映射到高维空间 2) 在高维空间构建模型

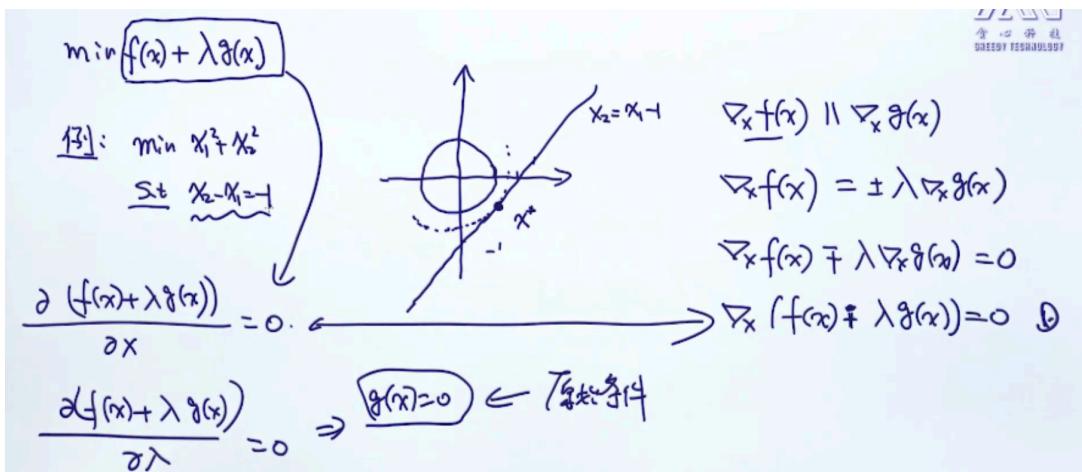


使用 kernel trick 把上述两步合成一步，降低了时间复杂度。

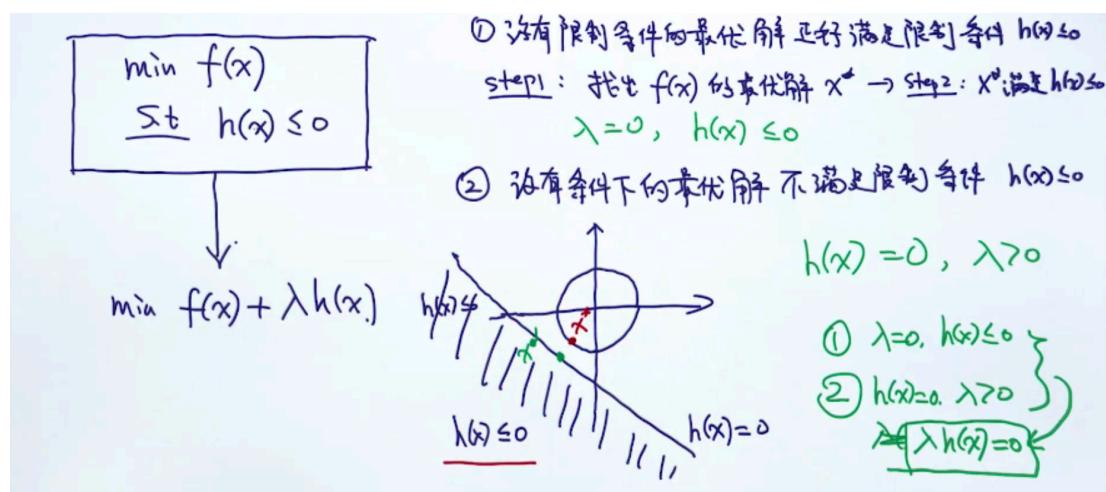
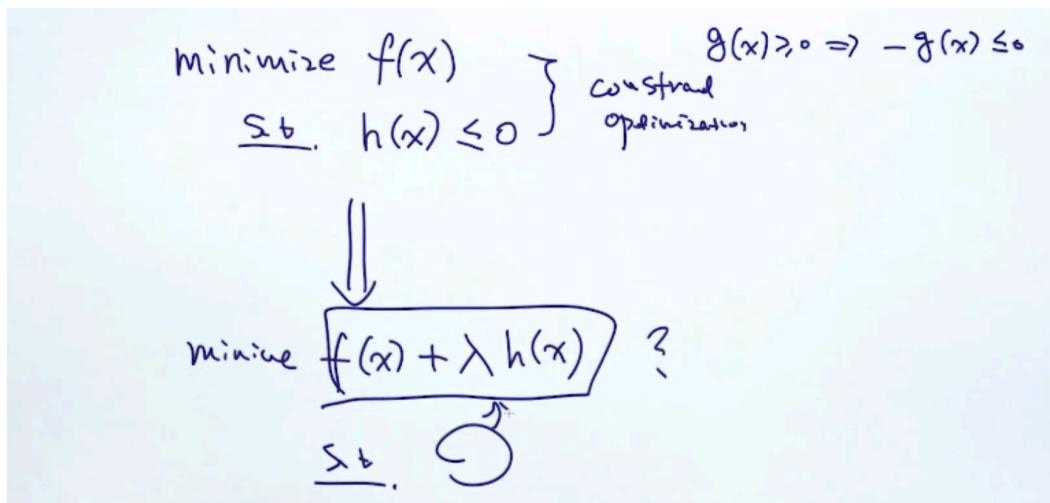
## 6. 拉格朗日优化条件（回顾一些优化的技巧）

### 6.1 等号条件

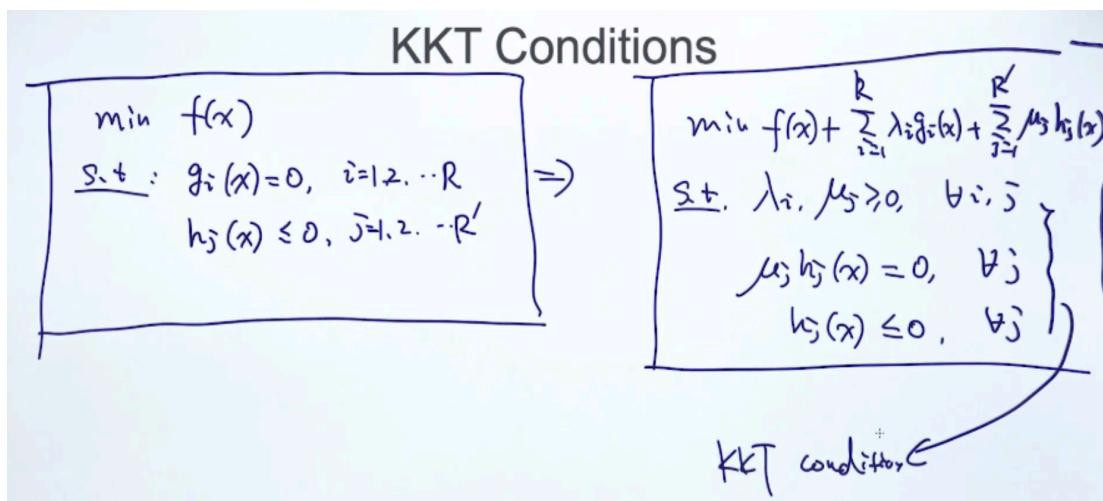




## 6.2 不等号条件



## 7. KKT 条件



## 7.1 SVM 的 KKT 条件

SVM - Hard Constraint

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|_2^2 \\ & \text{s.t. } y_i(w^T x_i + b) - 1 \geq 0 \quad \forall i \end{aligned}$$

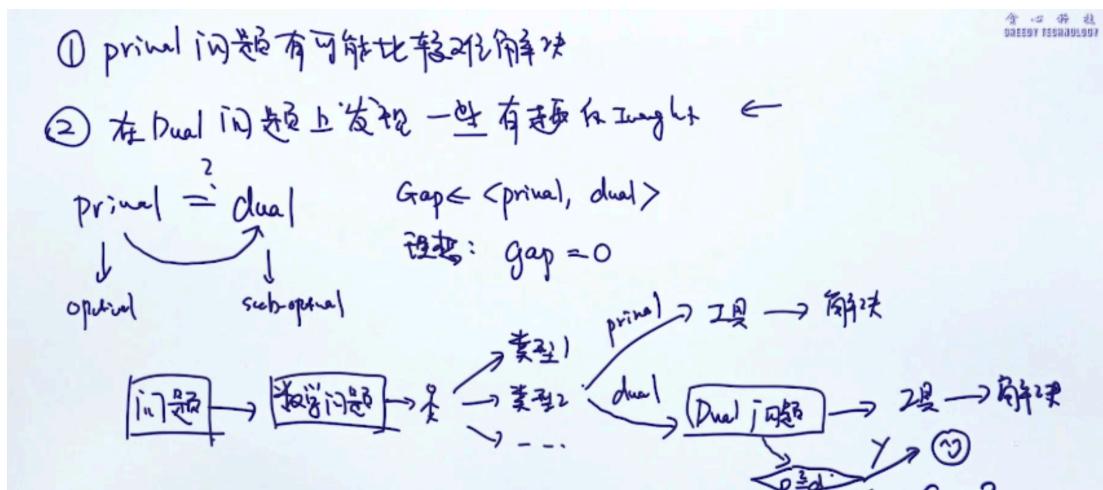
↓ 基本问题

primal problem

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|_2^2 + \sum_{i=1}^n \lambda_i [1 - y_i(w^T x_i + b)] \\ & \text{s.t. } \lambda_i \geq 0 \quad \forall i \\ & \quad \lambda_i [1 - y_i(w^T x_i + b)] = 0 \quad \forall i \\ & \quad 1 - y_i(w^T x_i + b) \leq 0 \quad \forall i \end{aligned}$$

SVM P.S.      KKT conditions

## 8. Primal-Dual



## Dual Derivation of SVM

$$L = \frac{1}{2} \|w\|_2^2 + \sum_{i=1}^n \lambda_i [1 - y_i(w^T x_i + b)]$$

S.t.  $\lambda_i \geq 0, \forall i$

$$\lambda_i [1 - y_i(w^T x_i + b)] = 0, \forall i$$

$$1 - y_i(w^T x_i + b) \leq 0, \forall i$$

$$\begin{aligned} \frac{1}{2} \|w\|_2^2 &= \frac{1}{2} w \cdot w = \frac{1}{2} \left( \sum_{i=1}^n \lambda_i y_i x_i \right)^T \left( \sum_{j=1}^n \lambda_j y_j x_j \right) \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i y_i x_i^T \lambda_j y_j x_j^T x_j \\ \sum_{i=1}^n \lambda_i [1 - y_i(w^T x_i + b)] &= \frac{n}{2} \lambda_1 - \sum_{i=1}^n \lambda_i y_i (\underbrace{\sum_{j=1}^n \lambda_j y_j x_j}_b) \\ &= \frac{n}{2} \lambda_1 - \sum_{i=1}^n \lambda_i y_i \underbrace{\lambda_1 y_1 x_1^T x_i}_b - \sum_{i=2}^n \lambda_i y_i \underbrace{\lambda_1 y_1 x_1^T x_i}_b \\ &= \frac{n}{2} \lambda_1 - \sum_{i=1}^n \lambda_i y_i (\sum_{j=1}^n \lambda_j y_j x_j)^T x_i \\ &= \frac{n}{2} \lambda_1 - \sum_{i=1}^n \lambda_i y_i (\sum_{j=1}^n \lambda_j y_j x_j)^T x_i \\ \therefore L &= -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i y_i x_i^T x_j + \frac{n}{2} \lambda_1 \end{aligned}$$

### Dual Formation of SVM

$$\text{minimize} \quad -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j x_i^T x_j + \frac{1}{2} \sum_{i=1}^n \lambda_i$$

S.t.  $\lambda_i \geq 0, \forall i$   $\Rightarrow$

$$\sum_{i=1}^n \lambda_i y_i = 0, \quad w = \sum_{i=1}^n \lambda_i y_i x_i$$

$$\lambda_i [1 - y_i(w^T x_i + b)] = 0$$

## 9. Kernel Trick

$$x \rightarrow \phi(x) = u$$

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \\ x_0 \end{pmatrix} \rightarrow \begin{matrix} \text{非线性} \\ \text{映射} \\ \text{函数} \end{matrix} \rightarrow u_1 \rightarrow u_2 \rightarrow \dots \rightarrow u_n$$

$D' \gg D$

$$O(D) \rightarrow O(D')$$

$$\phi(\phi(x)) = y$$

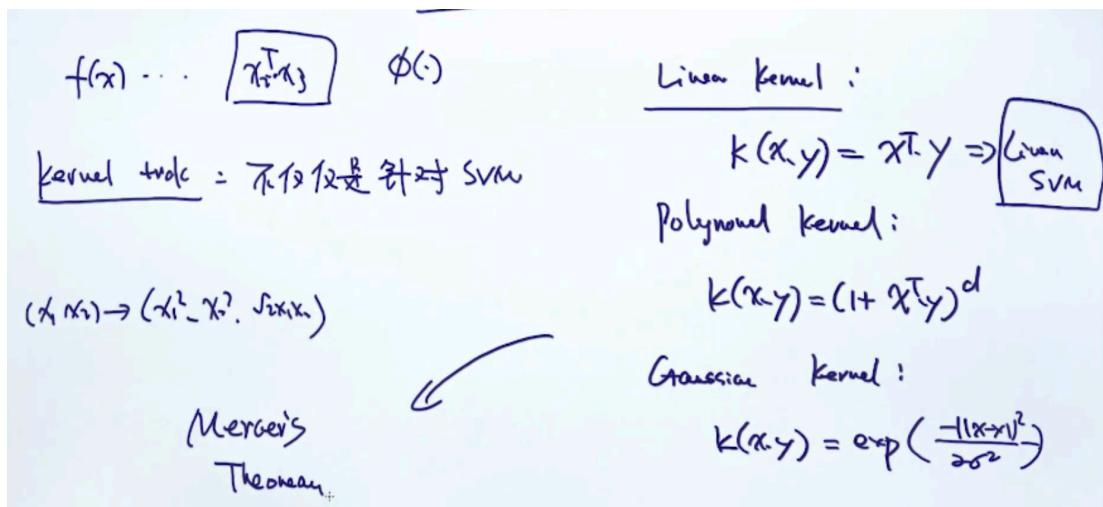
$$\phi(x) = (x_1, x_2), \quad z = (z_1, z_2)$$

$$x^T z = (x_1 z_1 + x_2 z_2)$$

$$\phi(x) = (x_1^2, x_2^2, \sqrt{2} x_1 x_2), \quad \phi(z) = (z_1^2, z_2^2, \sqrt{2} z_1 z_2)$$

$$\phi(x)^T \phi(z) = (x_1^2 z_1^2 + x_2^2 z_2^2 + 2 x_1 x_2 z_1 z_2)$$

$$= (x_1 z_1 + x_2 z_2)^2 = \boxed{(x^T z)^2}$$



## 10. 总结

- ① Linear SVM
  - Hard constraint
  - Soft - constraint (slack variable)
- ② KKT (KKT 条件)
  - 等式条件
  - 不等式条件
- ③ Dual Formulation (kernel trick)
  - primal  $\stackrel{?}{=}$  dual