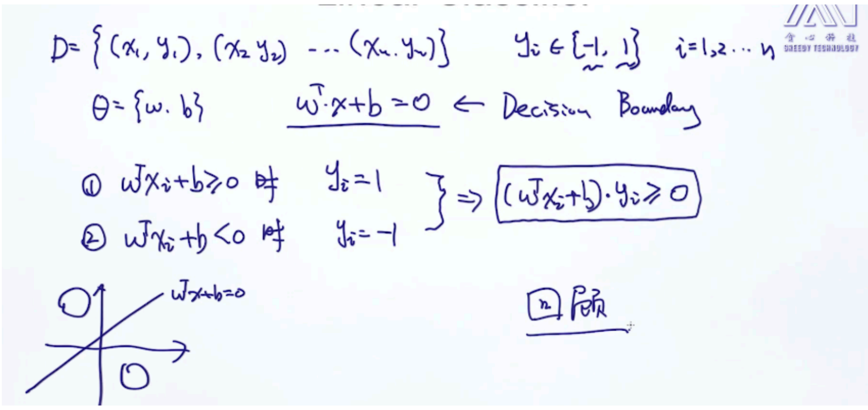
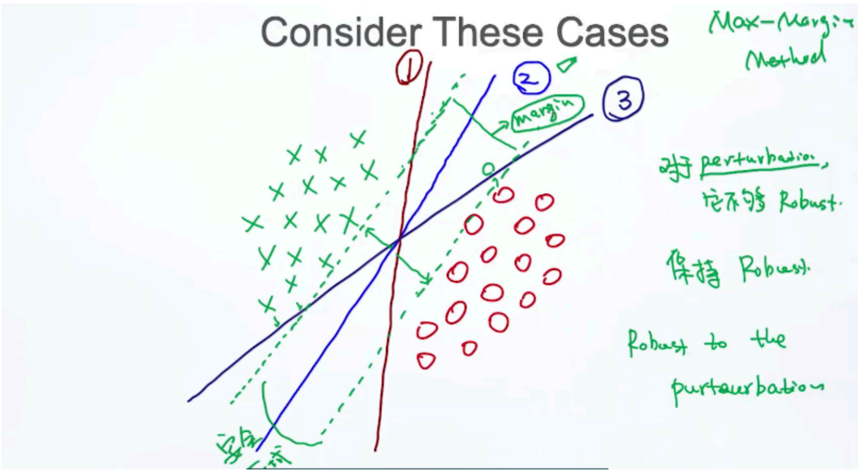
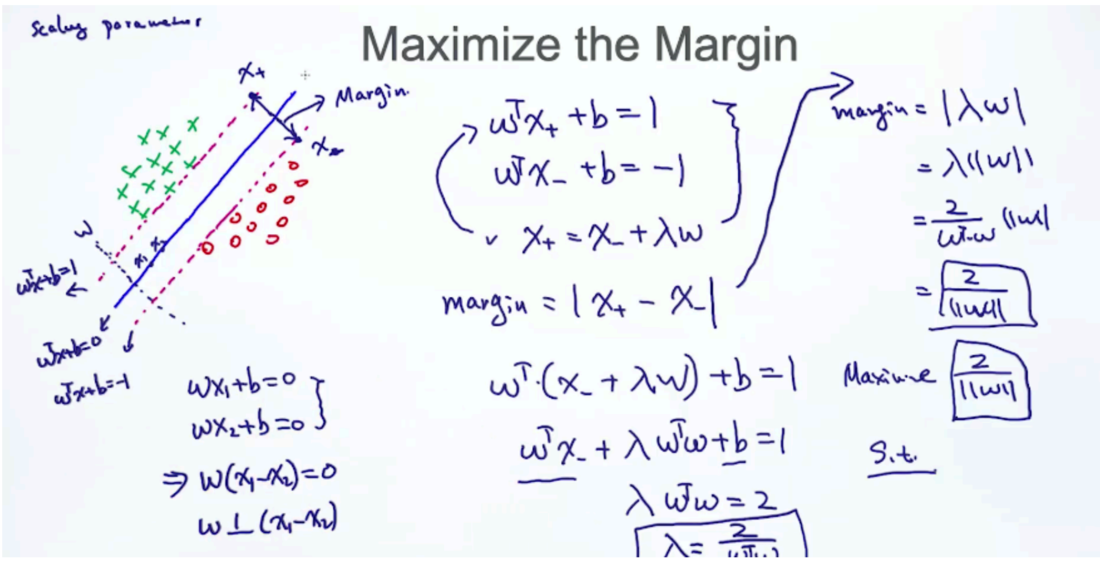
1. Linear Classifier



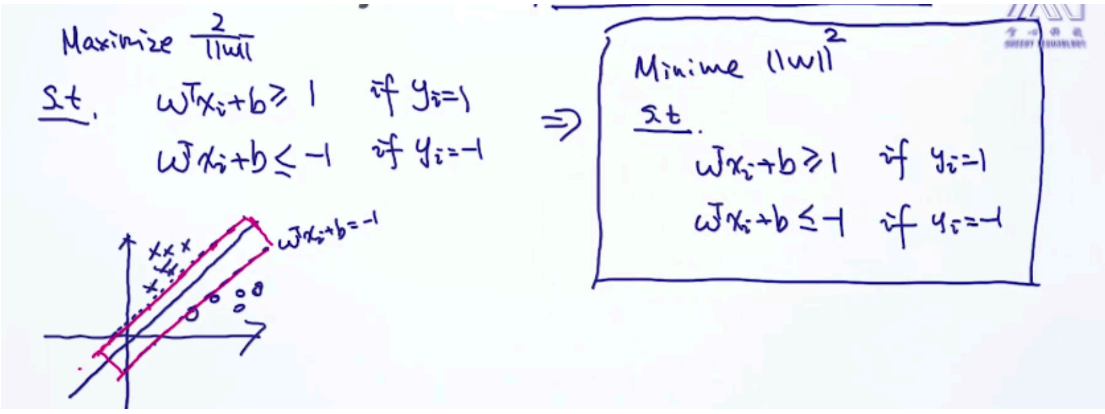


2. SVM

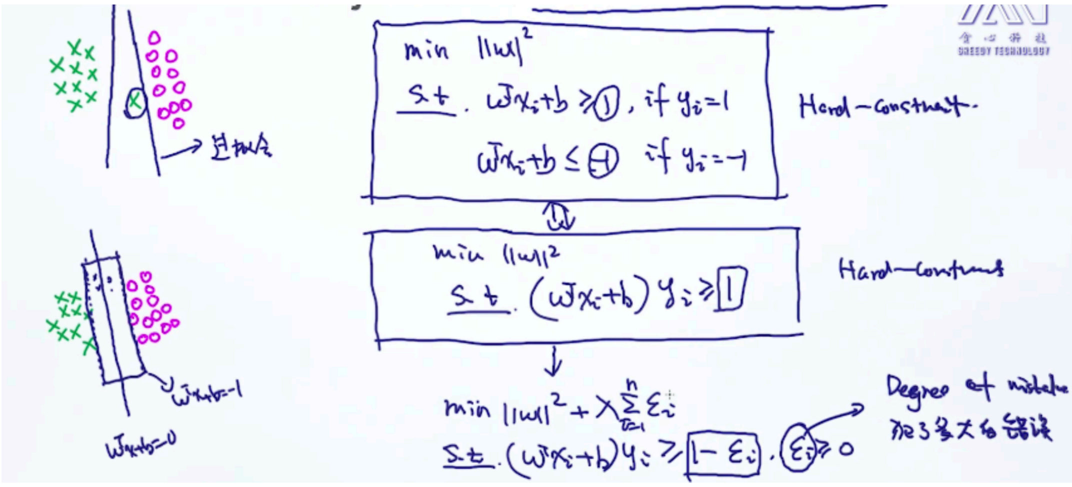
2.1 Maximize the margin



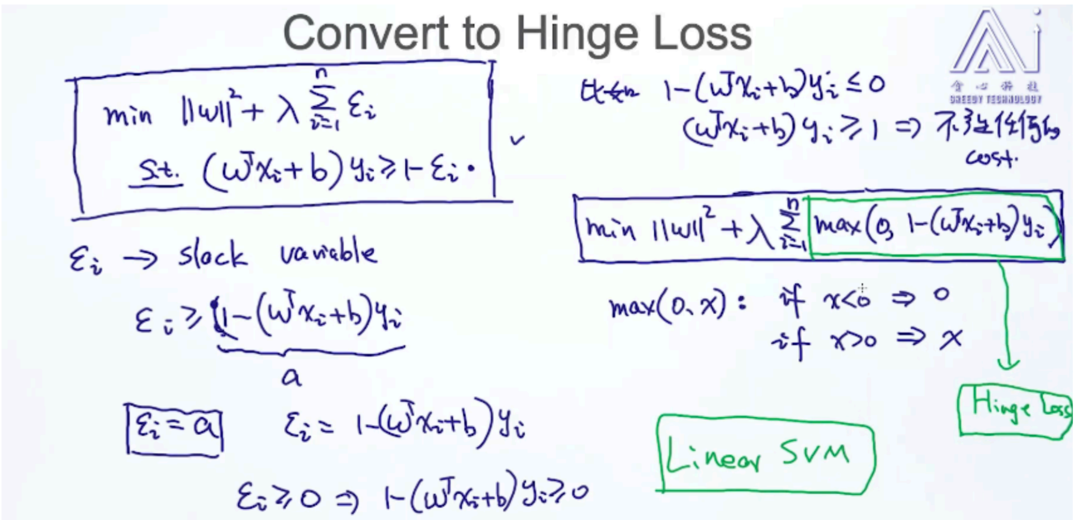
2.2 SVM 的目标函数:Hard constraint



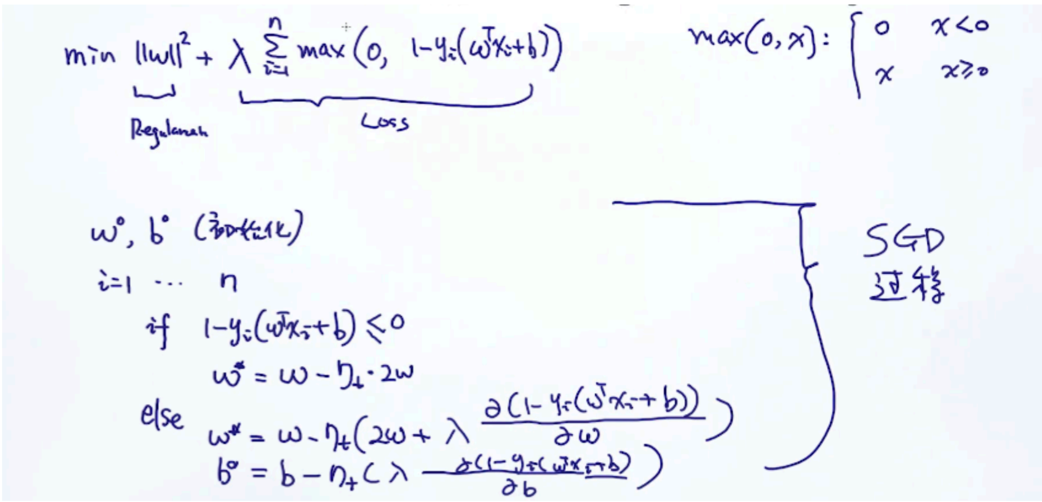
2.3 SVM 的目标函数:Soft constraint



3. Hinge Loss

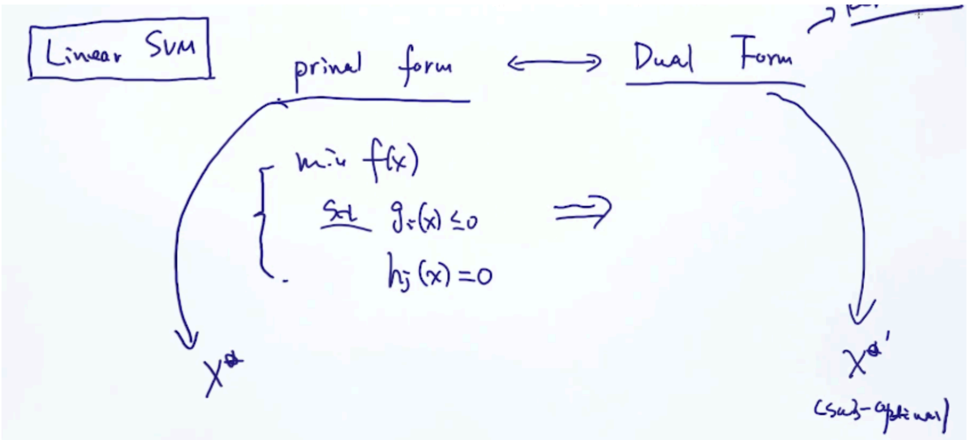


3.1 Gradient Descent for Hinge Loss Objective



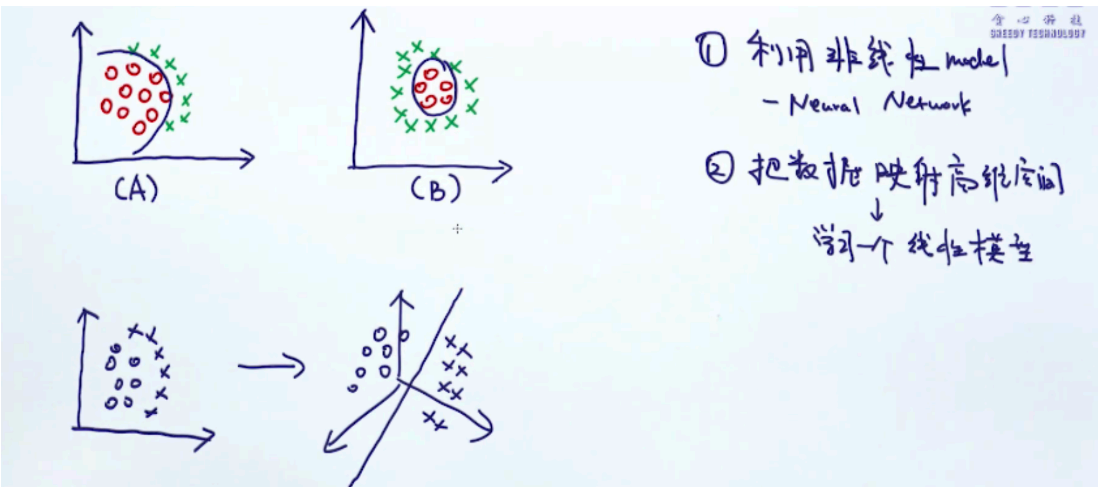
4. Linear SVM

4.1 Primal Form

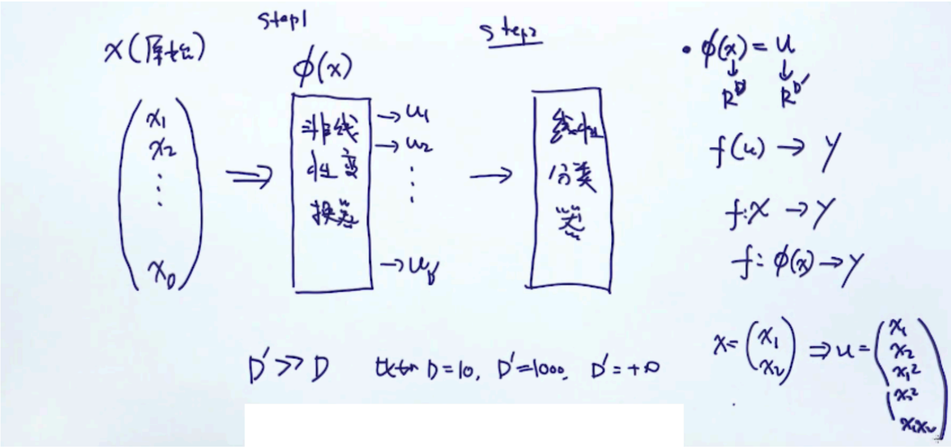


Kernel

4.2 linear SVM 的缺点

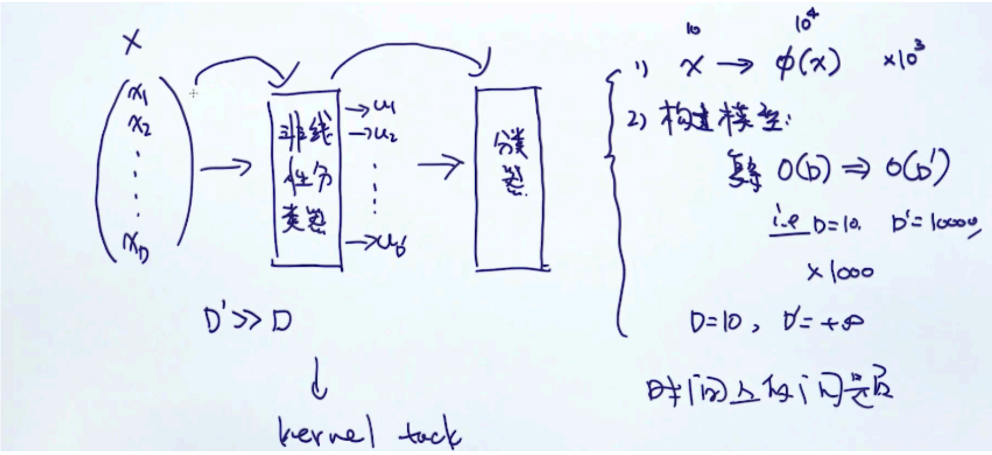


5. Mapping Feature to High Dimensional Space



缺点：时间复杂度提升了

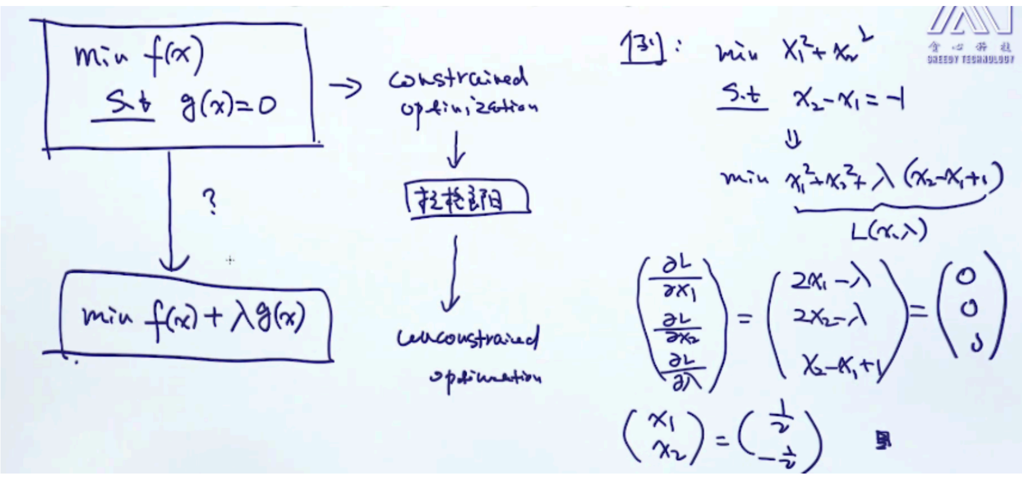
1）把原始特征映射到高维空间 2）在高维空间构建模型

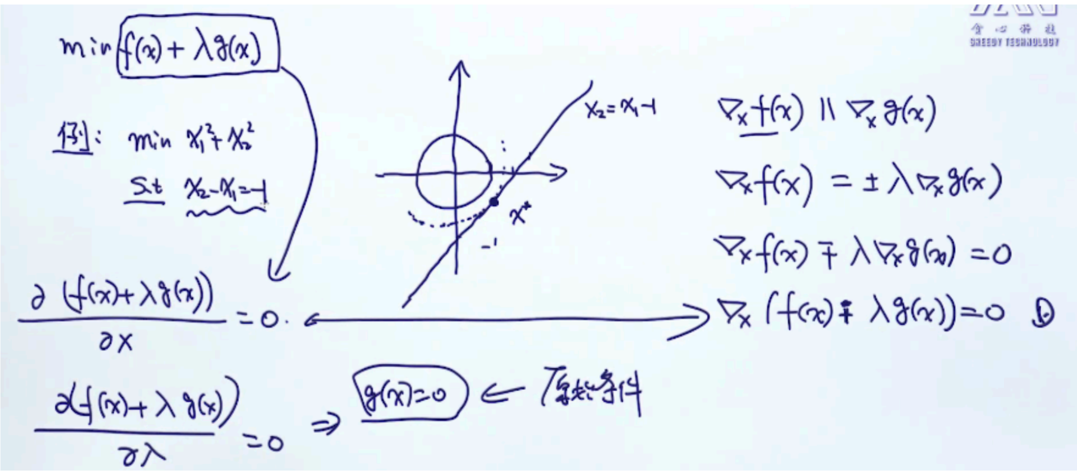


使用 kernel trick 把上述两步合成一步，降低了时间复杂度。

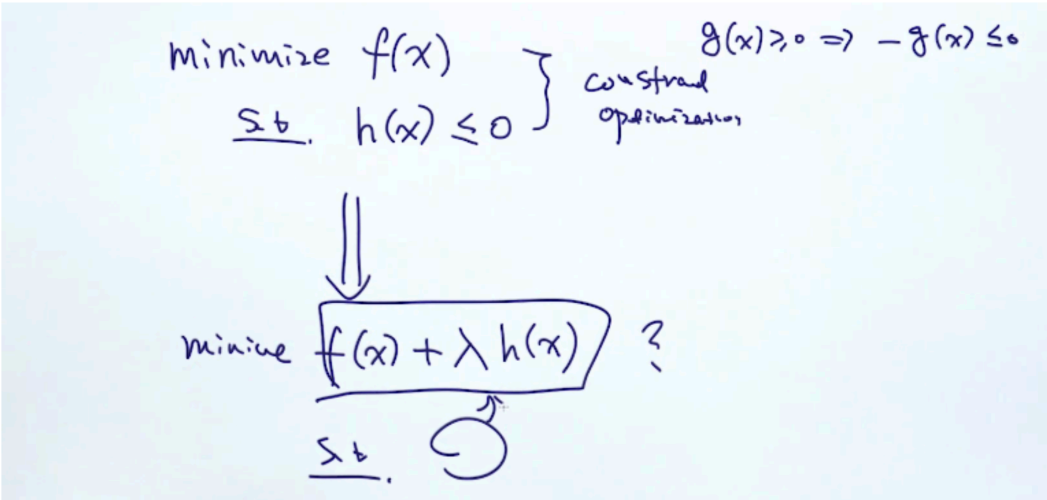
6. 拉格朗日优化条件（回顾一些优化的技巧）

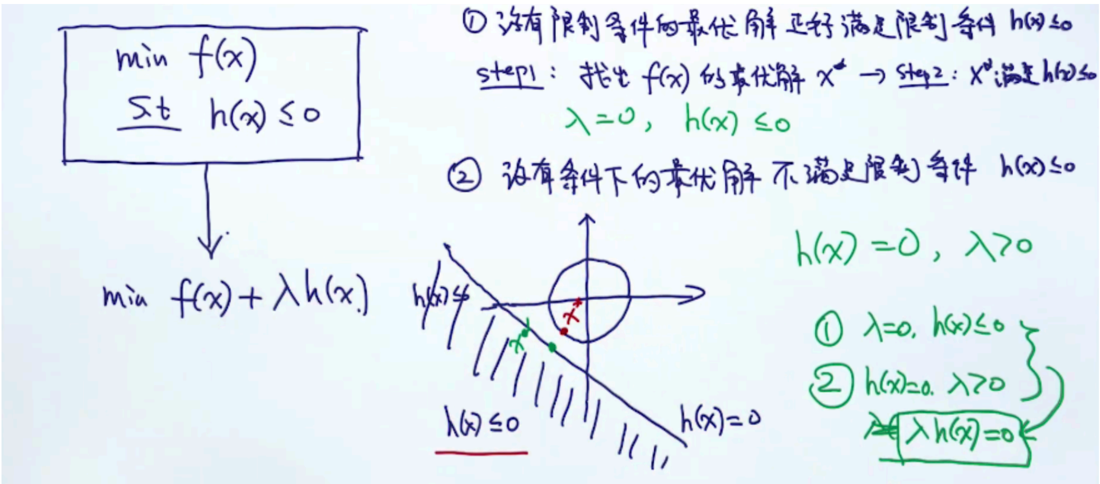
6.1 等号条件



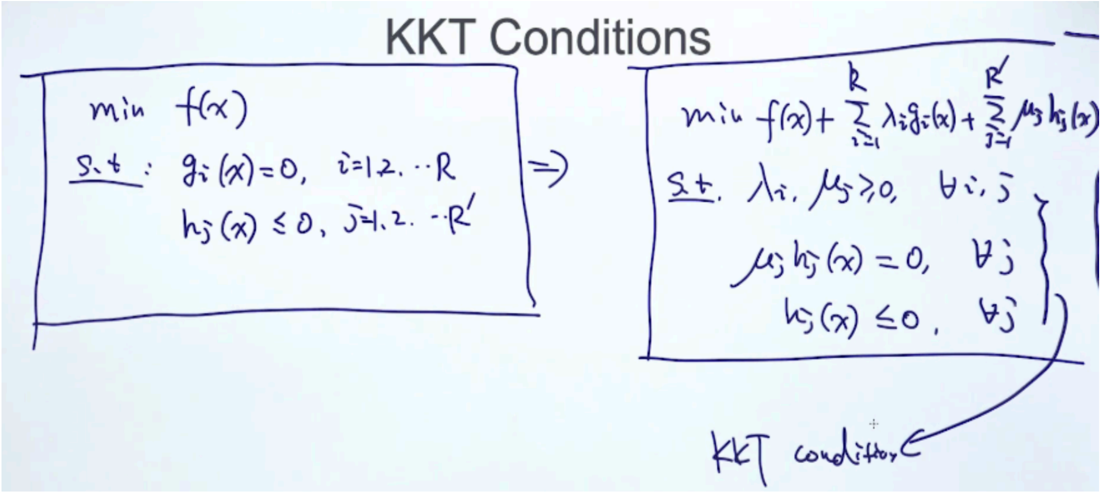


6.2 不等号条件

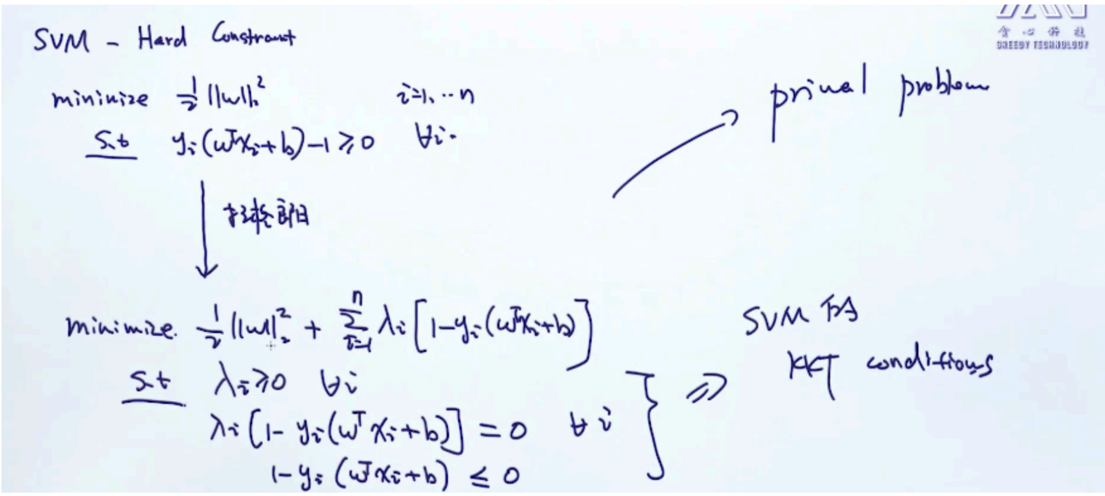




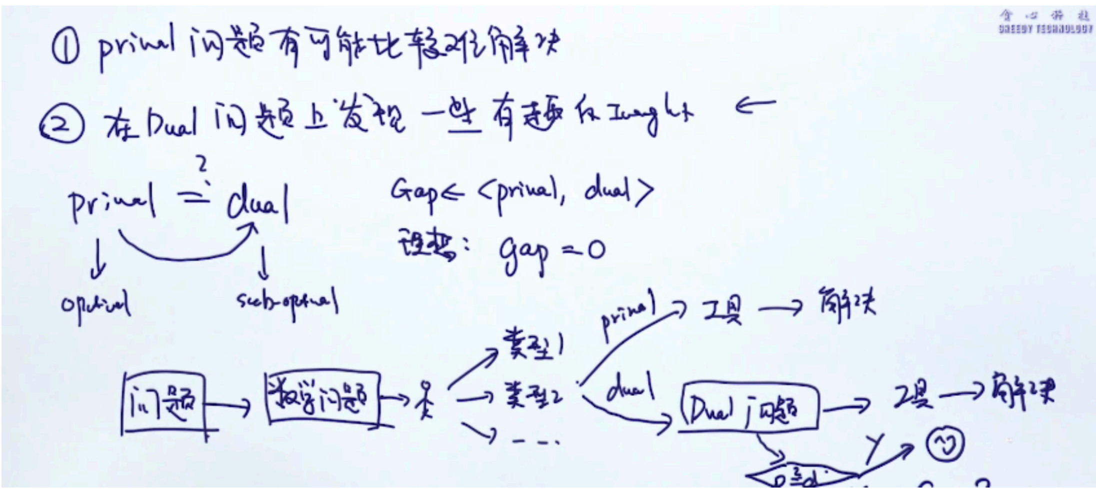
7. KKT 条件



7.1 SVM 的 KKT 条件



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. $\lambda_i \geq 0 \quad \forall i$

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

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

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w + \sum_{i=1}^{n} \lambda_i(-y_i x_i) \Rightarrow \boxed{w = \sum_{i=1}^{n} \lambda_i y_i x_i}$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \boxed{\sum_{i=1}^{n} \lambda_i y_i = 0}$$

$$\frac{1}{2}\|w\|_2^2 = \frac{1}{2} w^T \cdot w = \frac{1}{2}\left(\sum_{i=1}^{n} \lambda_i y_i x_i\right)^T \cdot \left(\sum_{j=1}^{n} \lambda_j y_j x_j\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j y_i y_j x_i^T \cdot x_j$$

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

$$= \sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{n} \lambda_i y_i w^T x_i - \cancel{\sum_{i=1}^{n} \lambda_i y_i b}$$

$$= \sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{n} \lambda_i y_i \left(\sum_{j=1}^{n} \lambda_j y_j x_j\right)^T \cdot x_i$$

$$= \sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j y_i y_j x_i^T \cdot x_j$$

$$\therefore \boxed{L = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j y_i y_j x_i^T \cdot x_j + \sum_{i=1}^{n} \lambda_i}$$

Dual

Dual Formation of SVM

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

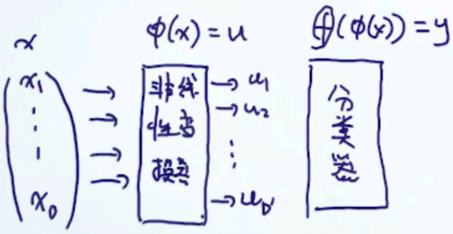
s.t. $\lambda_i \geq 0 \quad \forall i$

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

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

$\Rightarrow$ 未知参数 $\boxed{\lambda}$

9. Kernel Trick



$D' \gg D$

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

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

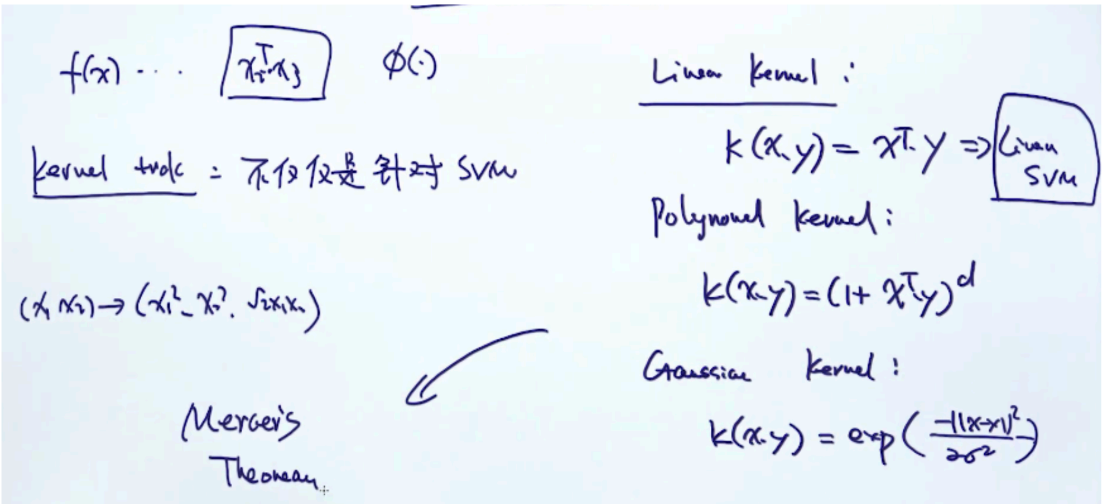
$\boxed{x^T \cdot z = (x_1 z_1 + x_2 z_2)}$

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

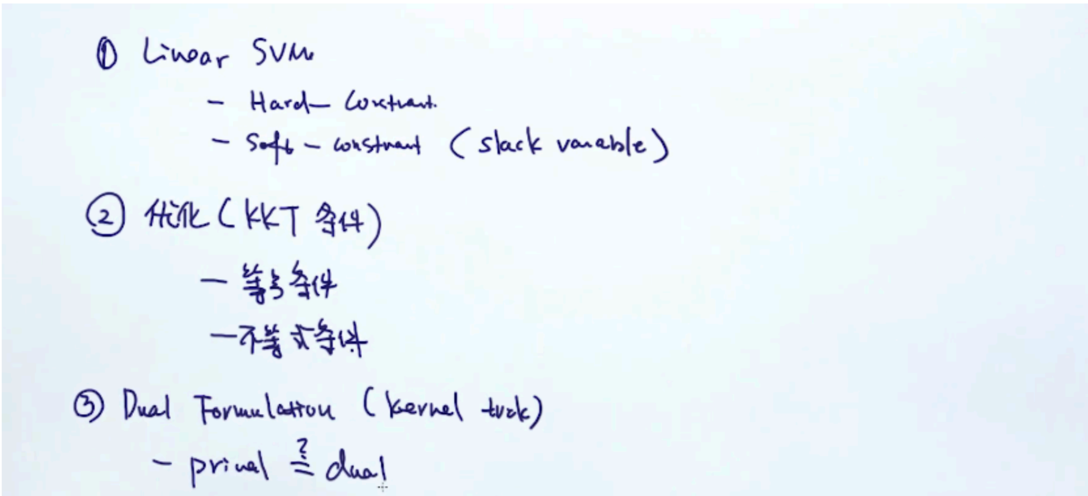
$\phi(x)^T \phi(z) = (x_1^2 z_1^2 + x_2^2 z_2^2 + 2 x_1 x_2 z_1 z_2)$

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

$(x^T \cdot z)^2$


$f(x) \cdots$ $\boxed{x_i^T x_j}$ $\phi(\cdot)$
Kernel trick : 不仅仅是针对 SVM
$(x_1, x_2) \rightarrow (x_1^2, x_2^2, \sqrt{2}x_1x_2)$
Mercer's Theorem
Linear kernel :
$K(x,y) = x^T y \Rightarrow$ Linear SVM
Polynomial kernel:
$K(x,y) = (1 + x^T y)^d$
Gaussian kernel :
$K(x,y) = \exp\left(\frac{-\|x-y\|^2}{2\sigma^2}\right)$

10. 总结


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