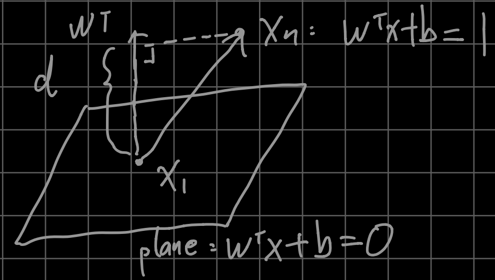
SVM

Derivation:

margin: $d = \frac{|w^T(x_n - x_1)|}{\|w\|} = \frac{1}{\|w\|}$



Intuition: maximize margin d.

$\Rightarrow$ Optimization problem:

$$\text{maximize } \frac{1}{\|w\|} \quad \text{s.t. } \min_{n=1,2\cdots N} |w^T x_n + b| = 1 \quad \Rightarrow \quad \text{minimize } \frac{1}{2} w^T w \quad \text{s.t. } y_i(w^T x_i + b) \geq 1$$

y_i: label, sample points 在 plane 上方 $y_i = 1$; 在下方 $y_i = 0$ ⇒ 为去 $|\cdot|$ 一种方法

Lagrangian Multiplier method:

$$\text{minimize } L(w, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{N} \alpha_i [y_i(w^T x_i + b) - 1]$$

constraints 为 '<' 则用 '+', '>' 则用 '-'

$$\begin{cases} \dfrac{\partial L(w,\alpha)}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{N} \alpha_i \cdot y_i \cdot x_i \quad \cdots (1) \\ \dfrac{\partial L(w,\alpha)}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0 \quad \cdots (2) \end{cases}$$

$\Downarrow$ 将 (1), (2) 代回 $L(w, \alpha)$ 得

Dual form:

$$\text{maximize } L(\alpha) = \frac{1}{2} \cdot \sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j x_i^T x_j$$

$$- \sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j x_i^T x_j - b \cdot \underbrace{\sum_{i=1}^{N} \alpha_i y_i}_{0} + \sum_{i=1}^{N} \alpha_i$$

$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j x_i^T x_j$$

$$\text{s.t.} \quad \alpha_i \geq 0, \quad i = 1, 2 \cdots N$$

$$\sum_{i=1}^{N} \alpha_i \cdot y_i = 0$$

$\Rightarrow$ Dual form 有 constraints，不要忘了。

We can found that dual form is a square form (二次型)。$\Rightarrow$ 有许多成熟的技术，e.g. SMO (Seqential Minimal Optimization) 等可解这个问题。

Solution:

$$\alpha_1, \alpha_2 \cdots \alpha_N$$

$$\Rightarrow \begin{cases} w = \sum_{i=1}^{N} \alpha_i y_i x_i \\ b = \frac{1}{y_i} - w^T x_i, \text{ if } y_i(w^T x_i + b) = 1 \end{cases}$$

(After get the hyperplane, we found the decision boundary) s.t. KKT.

$$\underbrace{\alpha_i \cdot (y_i(W^T X_i + b) - 1) = 0 \quad , i = 1, 2 \cdots N}_{\text{KKT equation (4)}}$$

KKT.

(1) $f_i(x) \le 0, \ i = 1, 2 \cdots m$

(2) $h_i(x) = 0, \ i = 1, 2 \cdots n$

(3) $\alpha_j \geq 0, \ j = 1, 2 \cdots m$

(4) $\alpha_i \cdot f_i(x) = 0, \ j = 1, 2 \cdots m$

(5) $\nabla f_0(x) - \sum_{i=1}^{m} u_i f_i(x) - \sum_{j=1}^{n} v_j \cdot h_j(x) = 0$

We can get:

① $\alpha_i > 0 \Rightarrow y_i(W^T X_i + b) = 1$

此时 point X_i 处于 hyperplane 上，是 support vector

② w, b 只与 support vector 有关。

Soft margin:

$$\text{minimize} \quad \frac{1}{2} W^T W - C \cdot \sum_{i=1}^{N} \xi_i$$

$$\text{s.t.} \quad y_i(W^T X_i + b) \geq 1 - \xi_i, \quad i = 1, 2 \cdots N$$

其中 ξ_i 为 tolerance，即容许一些点判错。C 为惩罚项。

Kernel SVM:

Use kernel trick: $K(x, x')$ to replace x^Tx.

$z = \phi(x)$，将 x 变换到高维的 z 后，使得原问题可解（线性可分）。而 Kernel trick 使得我们不用 explicitly 给出 $\phi(\cdot)$，只需直接给出 $z^T z$ 的表达式 $K(x, x')$ 即可。

Kernel Dual form:

$$L(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j K(x_i, x_j)$$

s.t. $\alpha_i \geq 0, \quad i = 1, 2 \cdots N$

$$\sum_{i=1}^{N} \alpha_i \cdot y_i = 0$$

⇒ Dual form 有 constraints，不要忘了。

求解 $\alpha_1, \alpha_2 \cdots \alpha_n$ 后用 inverse map 将 z 映射回 x，求得 w, b 为原问题的解。

GAN

Optimization object:

$$\min_G \max_D V(D,G) = \min_G \max_D E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim g(z)}[\log(1-D(G(z)))]$$

真实分布 sample 出来的用 x 表示

generator 产生的数据用 z 表示

Training process:

repeat

Optimize discriminator:

maximize $V(D,G)$ given G $\Rightarrow$ gradient ascent
D

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D(x_i) + \log(1 - D(G(z_i))) \right]$$

mini-batch

Optimize generator:

minimize $V(D,G)$ given D $\Rightarrow$ gradient descent
G

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(G(z_i)))$$

mini-batch

until converged.