1. Support Vector Machines Dual Problem - (40 pts). Assume that you are given a data set $\mathcal{D} = \{(t_i, \mathbf{x}_i) \ : \text{for } i = 1, ..., N\}$ with $t_i \in \{\pm 1\}$.

1.1. *Hard margin - (20 pts).* Recall that the hard margin SVM problem can be written in the following primal form

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|_2^2$$
$$\text{s.t. } t_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 \qquad i = 1, \ldots, N$$

(a) Write down the Lagrangian for this problem with Lagrangian parameters denoted with α_i's.

$$L(w, b, \alpha) = \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^{N} \alpha_i \left(1 - t_i(w^T x_i + b)\right)$$

(b) Show that the equivalent dual problem can be written as

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} t_i t_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j$$
$$\text{s.t. } 0 \leq \alpha_i \qquad i = 1, 2, ..., N.$$
$$\sum_{i=1}^{N} \alpha_i t_i = 0.$$

min :

① $\frac{\partial L}{\partial w} = w - \sum_{i=1}^{N} \alpha_i t_i x_i = 0.$

$\Rightarrow w^* = \sum_{i=1}^{N} \alpha_i t_i x_i$

② $\frac{\partial L}{\partial b} = 0 - \sum_{i=1}^{N} \alpha_i t_i = 0.$

$\Rightarrow \sum_{i=1}^{N} \alpha_i t_i = 0.$

$$L(w,b,\alpha) = \frac{1}{2}\left\|\sum_{i=1}^{N} \alpha_i t_i x_i\right\|_2^2 + \sum_{i=1}^{N} \alpha_i \left[1 - t_i\left(\left(\sum_{j=1}^{N} \alpha_j t_j x_j\right)^T x_i + b\right)\right]$$
$$= \sum_{i=1}^{N} \alpha_i + \frac{1}{2}\left\|\sum_{i=1}^{N} \alpha_i t_i x_i\right\|_2^2 - \sum_{i=1}^{N} \alpha_i t_i \left(\sum_{j=1}^{N} \alpha_j t_j x_j\right)^T x_i - \underbrace{\sum_{i=1}^{N} \alpha_i t_i b}_{=0}.$$
$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N} \alpha_i t_i \left(\sum_{j=1}^{N} \alpha_j t_j x_j\right)^T x_i$$
$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} t_i t_j \alpha_i \alpha_j \underbrace{x_j^T x_i}_{= x_i^T x_j}$$
$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} t_i t_j \alpha_i \alpha_j x_i^T x_j = W(\alpha)$$

we should minimize $L(w,b,\alpha)$, which is the same as min $W(\alpha)$

Also, from before, $\sum_{i=1}^{N} \alpha_i t_i = 0$. , $\alpha_i \geq 0 \ \forall i$ (Assumption, otherwise α_i can be $-\infty$ to minimize)

$\Rightarrow$ the problem is the same as dual problem

(c) Assume that we solved the above dual formulation and obtained the optimal α. For a given test data point $\mathbf{x}$, how can we predict its class?

$$\hat{y} = w^{*T} x + b^*$$
$$= \left(\sum_{i=1}^{N} \alpha_i t_i x_i\right)^T x + b^*$$

if : $\hat{y} > 0$: class +1

if : $\hat{y} < 0$: class −1

solve b^* by:

find x_1, x_2 on two margins

solve the equations : $(t_1 = -t_2)$

$$\begin{cases} w^{*T} x_1 + b = t_1 \\ w^{*T} x_2 + b = t_2 \end{cases} \Rightarrow b^* = -\frac{1}{2} w^T (x_1 + x_2)$$

1.2. *Soft margin - (20 pts).* Recall that the soft margin SVM problem can be written in the following primal form

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}\|\mathbf{w}\|_2^2 + \gamma \sum_{i=1}^{N} \xi_i$$
$$\text{s.t.} \quad t_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i \qquad i = 1, \ldots, N$$
$$\xi_i \geq 0 \qquad i = 1, \ldots, N$$

(a) Use the Lagrangian provided in the lecture to show that the equivalent dual problem can be written as

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} t_i t_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j$$
$$\text{s.t.} \quad 0 \leq \alpha_i \leq \gamma \qquad i = 1, \ldots, N$$
$$\sum_{i=1}^{N} \alpha_i t_i = 0.$$

$$L(w, b, \alpha, \xi, \gamma) = \frac{1}{2}\|w\|_2^2 + \gamma \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i [t_i(w^T x_i + b) - 1 + \xi_i] - \sum_{i=1}^{N} \beta_i \xi_i$$

min :

① $$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{N} \alpha_i t_i x_i$$

$$\Rightarrow w^* = \sum_{i=1}^{N} \alpha_i t_i x_i$$

② $$\frac{\partial L}{\partial b} = -\sum_{i=1}^{N} \alpha_i t_i = 0$$

$$\Rightarrow \sum_{i=1}^{N} \alpha_i t_i = 0$$

③ $$\frac{\partial L}{\partial \xi_i} = \gamma - \alpha_i - \beta_i = 0$$

$$\Rightarrow \gamma = \alpha_i + \beta_i$$

$$L(w, b, \alpha, \xi, \gamma) = \frac{1}{2} \left\| \sum_{i=1}^{N} \alpha_i t_i x_i \right\|_2^2 + \cancel{\sum_{i=1}^{N} (\alpha_i + \beta_i) \xi_i}$$

$$- \sum_{i=1}^{N} \alpha_i \left[t_i \left(\left(\sum_{j=1}^{N} \alpha_j t_j x_j \right)^T x_i + b \right) - 1 \right) - \cancel{\sum_{i=1}^{N} \alpha_i \xi_i - \sum_{i=1}^{N} \beta_i \xi_i}$$

$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \left\| \sum_{i=1}^{N} \alpha_i t_i x_i \right\|_2^2 - \sum_{i=1}^{N} \alpha_i t_i \left(\sum_{j=1}^{N} \alpha_j t_j x_j \right)^T x_i - \underbrace{\sum_{i=1}^{N} \alpha_i t_i b}_{0.}$$

$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} t_i t_j \alpha_i \alpha_j x_i^T x_j \quad = W(\alpha)$$

min $L(w, b, \alpha, \xi, \gamma)$ is same as min $W(\alpha)$

s.t. $\sum_{i=1}^{N} \alpha_i t_i = 0$ (from before), $\alpha_i, \beta_i \geq 0 \Rightarrow 0 \leq \alpha_i \leq \gamma \quad \forall i.$

(b) Assume that we solved the above dual formulation and obtained the optimal α. For a given test data point $\mathbf{x}$, how can we predict its class?

$$\hat{y} = w^{*T} x + b^*$$

$$= \left(\sum_{i=1}^{N} \alpha_i t_i x_i \right)^T x + b^*$$

if $\hat{y} > 0$ then class $+1$

if $\hat{y} < 0$ then class -1

solve b^* by:

find x_1, x_2 on two margins

solve the equations: $(t_1 = -t_2)$

$$\begin{cases} w^{*T} x_1 + b = t_1 \\ w^{*T} x_2 + b = t_2 \end{cases} \Rightarrow b^* = -\frac{1}{2} w^T (x_1 + x_2)$$