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.

The lagrangian for this problem becomes

$L(w,b,\alpha) = \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^N \alpha_i[t_i(w^T x_i + b) - 1]$ and $\alpha_i \geq 0 \ \forall \alpha$

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

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

Using previous lagrangian problem we have

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

The goal of the lagrangian is to solve for,

$\min_{w,b} \max_\alpha L(w,b,\alpha)$

or equivantley,

$\max_\alpha \min_{w,b} L(w,b,\alpha)$

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

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

Let it equal to 0 we have

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

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

Let it equal to 0 we have

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

Substituting these back into Lagrangian,

(Dual): $\max_\alpha \frac{1}{2}\|\sum_{i=1}^N \alpha_i t_i x_i\|_2^2 - \sum_{i=1}^N \alpha_i[t_i(\sum \alpha_i t_i x_i^2 + b) - 1]$

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

This is 0 (constraint)

$$= \max_{\alpha} \frac{1}{2}\sum_{i,j=1}^{N} t_i t_j \alpha_i \alpha_j x_i^T x_j - \sum_{i,j=1}^{N} t_i t_j \alpha_i \alpha_j x_i^T x_j + \sum_{i=1}^{N} \alpha_i$$

$$= \max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} t_i t_j \alpha_i \alpha_j x_i^T x_j$$

Here we have 1 more constraints

$$\alpha_i \geq 0 \ , \quad \sum_{i=1}^{N} \alpha_i t_i = 0$$

$$\boxed{\text{So (Dual)} = \max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} t_i t_j \alpha_i \alpha_j x_i^T x_j \quad \text{s.t } \alpha_i \geq 0, \ \sum_{i=1}^{N} \alpha_i t_i = 0}$$

(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?

Assume optimal α is given from above dual formulation

Then we have $w = \sum_{i=1}^{N} \alpha_i t_i x_i$ as previously shown.

Using the primal formulation :

$$t_i (w^T x_i + b) = 1$$

$$t_i^2 (w^T x_i + b) = t_i$$

$$w^T x_i + b = t_i$$

$$b = t_i - w^T x_i$$

So, with optimal α, we have w and b.

Using our classifier $f(x) = \text{sign}(w^T x + b)$

we can predict the class of a new test data x

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

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

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

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

The Lagrangian is given as the following:

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

where α_i, β_i are lagrange multipliers, constrained to be ≥ 0

① Setting derivative with respect to w to be 0:

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

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

② Setting derivative with respect to b to be 0:

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

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

③ Setting derivative with respect to ξ to be 0:

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

so $\gamma = \alpha_i + \beta_i$

Since $\beta_i \geq 0$ because Lagrange multiplier

we have $\gamma - \alpha_i = \beta_i \geq 0$

so $\alpha_i \leq \gamma$

Then $0 \leq \alpha_i \leq \gamma$ since $0 \leq \alpha_i$ α_i is lagrange multiplier.

Substitute w and Evaluate :

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

$$L(w,b,\xi,\alpha,\gamma) = \frac{1}{2}\sum_{i,j=1}^{N} t_i t_j \alpha_i \alpha_j x_i^T x_j + \gamma\sum_{i=1}^{N}\xi_i$$
$$- \sum_{i=1}^{N}\alpha_i t_i\left(\sum_{j=1}^{N}\alpha_j t_j x_j\right)x_i - b\sum_{i=1}^{N}\alpha_i t_i + \sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\alpha_i\xi_i - \sum_{i=1}^{N}\beta_i\xi_i$$

$$L(w,b,\xi,\alpha,\gamma) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} t_i t_j \alpha_i \alpha_j x_i^T x_j + \gamma\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\xi_i - \sum_{i=1}^{N}\beta_i\xi_i$$

$$= \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} t_i t_j \alpha_i \alpha_j x_i^T x_j + \sum_{i=1}^{N}(\gamma - \alpha_i - \beta_i)\xi_i$$

$$\boxed{\max_{\alpha} W(\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} t_i t_j \alpha_i \alpha_j x_i^T x_j \quad \text{such that } 0 \le \alpha_i \le \gamma \quad \forall \alpha_i, \quad \sum_{i=1}^{N}\alpha_i t_i = 0}$$

■

(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?

Assume optimal α is given from above dual formulation

Then we have $w = \sum_{i=1}^{N}\alpha_i t_i x_i$ as previously shown.

Using the primal formulation :

$$t_i(w^T x_i + b) = 1$$
$$t_i^2(w^T x_i + b) = t_i$$
$$w^T x_i + b = t_i$$
$$b = t_i - w^T x_i$$

So, with optimal α, we have w and b.

Using our classifier $f(x) = \text{sign}(w^T x + b)$

We can predict the class of a new test data x

■