SUPPORT VECTOR MACHINES CONTINUED.

1. SVM for Overlapping Classes

- In the previous case, $t_n\left(\mathbf{w}^t\phi(\mathbf{x}_n)+b\right)\geq 1$ is a constraint. It is required. What if the classes are not linearly separable in the kernel space? No solution may exist. So, we must allow for misclassification.
- We do this through slack variables, ξ
- The slack varaibles are $\xi_n \geq 0$
- If $\xi_n = 0$ if correctly classified on the boundary or farther away
- $\xi_n = |t_n - y(\mathbf{x}_n)|$ for other points. Increases as the data point is more mis-classified.
- So, the new constraints are:

$$t_n y(\mathbf{x}_n) \geq 1-\xi_n \quad \text{where } \xi_n \geq 0 \tag{1}$$

- So the objective function become:

$$C\sum_{n=1}^{N}\xi_n + \frac{1}{2}\|\mathbf{w}\|^2 \tag{2}$$

where $C>0$ is a trade-off parameter
- So, our Lagrangian becomes:

$$\mathscr{L}(\mathbf{w},b,\mathbf{a}) = C\sum_{n=1}^{N}\xi_n + \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N}a_n\left\{t_n y(x_n) - 1 + \xi_n\right\} - \sum_{n=1}^{N}\mu_n\xi_n \tag{3}$$

where $a_n \geq 0$, $\mu_n \geq 0$ are Lagrange multipliers
- *What is the meaning of each term?*
- *So, what are the resulting KKT conditions?*
- We now need to optimize with respect to $\mathbf{w}$, b, and ξ_n. How do we do this?

$$\tilde{\mathscr{L}}(\mathbf{a}) = \sum_{n=1}^{N}a_n - \frac{1}{2}\sum_n\sum_m a_n a_m t_n t_m K(\mathbf{x}_n,\mathbf{x}_m) \tag{4}$$

where $0 \leq a_n \leq C$ and $\sum_n a_n t_n = 0$
- The constraints are:

$$t_n y(\mathbf{x}_n) \geq 1-\xi_n \quad \text{where } \xi_n \geq 0 \tag{5}$$

- So the objective function becomes:

$$C\sum_{n=1}^{N}\xi_n + \frac{1}{2}\|\mathbf{w}\|^2 \tag{6}$$

where $C>0$ is a trade-off parameter

- Our Lagrangian is then:

$$\mathscr{L}(\mathbf{w}, b, \mathbf{a}) = C\sum_{n=1}^{N}\xi_n + \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n\left\{t_n y(x_n) - 1 + \xi_n\right\} - \sum_{n=1}^{N}\mu_n\xi_n \tag{7}$$

where $a_n \geq 0$, $\mu_n \geq 0$ are Lagrange multipliers
- *What is the meaning of each term?*
- *So, what are the resulting KKT conditions?*

$$a_n \geq 0 \tag{8}$$
$$t_n y(x_n) - 1 + \xi_n \geq 0 \tag{9}$$
$$a_n(t_n y(x_n) - 1 + \xi_n) = 0 \tag{10}$$
$$\xi_n \geq 0 \tag{11}$$
$$\mu_n \geq 0 \tag{12}$$
$$\mu_n\xi_n = 0 \tag{13}$$

- We now need to optimize with respect to $\mathbf{w}$, b, and ξ_n. How do we do this?

$$\frac{\partial\mathscr{L}}{\partial\mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_{n=1}^{N} a_n t_n \phi(\mathbf{x}_n) \tag{14}$$
$$\frac{\partial\mathscr{L}}{\partial\mathbf{b}} = 0 \rightarrow \sum_{n=1}^{N} a_n t_n = 0 \tag{15}$$
$$\frac{\partial\mathscr{L}}{\partial\xi} = 0 \rightarrow a_n = C - \mu_n \tag{16}$$

- Plug these into the Lagrangian to get the Dual form:

$$\begin{aligned}
\mathscr{L}(\mathbf{w}, b, \mathbf{a}) &= C\sum_{n=1}^{N}\xi_n + \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n\{t_n y(x_n) - 1 + \xi_n\} - \sum_{n=1}^{N}\mu_n\xi_n \qquad (17)\\
&= \sum_{n=1}^{N}\xi_n(C-\mu_n) + \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n\left\{t_n\left(\mathbf{w}^T\phi(\mathbf{x}_n) + b\right) - 1 + \xi_n\right\} \qquad (18)\\
&= \sum_{n=1}^{N}\xi_n a_n + \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n\left\{t_n\left(\mathbf{w}^T\phi(\mathbf{x}_n)\right)\right\} - b\sum_{n=1}^{N}a_n t_n + \sum_{n=1}^{N}a_n - \sum_{n=1}^{N}a_n\xi_n \qquad (19)\\
&= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n\left\{t_n\left(\mathbf{w}^T\phi(\mathbf{x}_n)\right)\right\} + \sum_{n=1}^{N}a_n \qquad (20)\\
&= \sum_{n=1}^{N}a_n + \frac{1}{2}\left(\sum_{n=1}^{N}a_n t_n\phi(\mathbf{x}_n)\right)^T\left(\sum_{n=1}^{N}a_n t_n\phi(\mathbf{x}_n)\right) \qquad (21)\\
&\quad - \sum_{n=1}^{N}a_n\left\{t_n\left(\left(\sum_{m=1}^{N}a_m t_m\phi(\mathbf{x}_m)\right)^T\phi(\mathbf{x}_n)\right)\right\} \qquad (22)\\
&= \sum_{n=1}^{N}a_n - \frac{1}{2}\sum_{n}\sum_{m}a_n a_m t_n t_m K(\mathbf{x}_n, \mathbf{x}_m) \qquad (23)
\end{aligned}$$

$$\tilde{\mathscr{L}}(\mathbf{a}) = \sum_{n=1}^{N}a_n - \frac{1}{2}\sum_{n}\sum_{m}a_n a_m t_n t_m K(\mathbf{x}_n, \mathbf{x}_m) \tag{24}$$

where $0 \le a_n \le C$ and $\sum_n a_n t_n = 0$

- *Why did we want it in this form?*