OBJECTIVE function for Soft - Margin SVM

$$\underset{\{w, b\}}{\arg\min} \ \frac{1}{2} \|w\|^2 + C \sum_{n=1}^{N} \varepsilon_n$$

subject to $t_n \cdot y(x_n) \geq 1 - \varepsilon_n, \ \forall n$ (1)

and $\varepsilon_n \geq 0, \ \forall n$ (2)

C = hyperparameter, $C > 0$

Primal Lagrangian.

$$\mathcal{L}(w, b, \varepsilon, a, \mu) = \frac{1}{2}\|w\|^2 + c \cdot \sum_{n=1}^{N} \varepsilon_n - \sum_{n=1}^{N} a_n \cdot \left(t_n \cdot y(x_n) - 1 + \varepsilon_n\right) - \sum_{n=1}^{N} \mu_n \cdot \varepsilon_n$$

Lagrange multipliers for constraint 1

Lagrange multipliers for constraint 2

where $y(x_n) = w^T \phi(x_n) + b$

KKT conditions:

$$\begin{cases} a_n \geq 0 \\ t_n \cdot y(x_n) - 1 + \varepsilon_n \geq 0 \\ a_n (t_n y(x_n) - 1 + \varepsilon_n) = 0 \\ \mu_n \geq 0 \\ \varepsilon_n \geq 0 \\ \mu_n \cdot \varepsilon_n = 0 \end{cases}$$

these are associated w/ 1

these are associated w/ 2

Optimizing for $\mathcal{L}$:

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial w} = 0 \\ \dfrac{\partial \mathcal{L}}{\partial b} = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \varepsilon_n} = 0 \end{cases} \iff \begin{cases} w = \sum_{n=1}^{N} a_n . t_n . \phi(x_n) \\ \sum_{n=1}^{N} a_n . t_n = 0 \\ c - a_n - \mu_n = 0 \iff a_n = c - \mu_n \end{cases}$$

$\Rightarrow a_n \leq c$

because $\mu_n \geq 0$
and $c > 0$

Dual Lagrangian:

Plugging in the solution for w in $\mathcal{L}$:

$$\tilde{\mathcal{L}}(a,\mu) = \sum_{n=1}^{N} a_n - \sum_{n=1}^{N}\sum_{m=1}^{N} a_n . a_m . t_n . t_m . K(x_n, x_m)$$

Subject to $\sum_{n=1}^{N} a_n . t_n = 0$

and $0 \leq a_n \leq C$

TAKES the form of a QUADRATIC PROGRAMMING PROBLEM ← USE AN ALGORITHM to find solution for a and μ.

① If $a_n = 0$ then $t_n . y(x_n) > 1 - \mathcal{E}_n$, which means that the point x_n is not a S.V.

② If $a_n > 0$ then $t_n . y(x_n) = 1 - \mathcal{E}_n$

(2.1) If $a_n < C$, then $\mu_n > 0$ because $a_n = C - \mu_n$.
From KKT conditions, if $\mu_n > 0$ then $\mathcal{E}_n = 0$.

(2.2) If $a_n = C$, then all corresponding points x_n will have $\mathcal{E}_n > 0$ $\Longrightarrow$ which means they are misclassified (wrong side of discriminant fct) or inside the margin (correct side of disc. fct).

③ If x_n is inside the margin (correct side of discriminant fct),

then $0 < \xi_n < 1$.

④ Otherwise, if they are misclassified,

then $\xi_n \geqslant 1$.

Discriminant function:

$$y(x) = w^T \phi(x) + b$$

$$\Leftrightarrow \boxed{y(x) = \sum_{n=1}^{N} a_n . t_n . K(x_n, x) + b}$$

$a_n \neq 0$ for all x_n that are:

① S.V.s

② correctly classified but inside the margin

③ misclassified point.

To solve for b, we know that the S.V.s

x_n, satisfy : $t_n \cdot y(x_n) = 1$ $(\xi_n = 0)$

Replacing $y(x_n)$ in this eq., we can solve for b:

$$\boxed{b = \frac{1}{N_M} \sum_{n \in M} \left(t_n - \sum_{m \in S} a_m \cdot t_m \cdot k(x_n, x_m) \right)}$$

$N_M \equiv$ cardinality of set of points that have $0 < a_n < C$. ("stable" points)

$S \equiv$ set of the S.V.s