$Hw4\ Shifeng\ Li \quad U24567277$

Bishop 6.2

The weight vector can be represent as:

$w = \sum_{n=1}^{N} \alpha_n t_n \phi(x_n)$

where n is α_n integer specifying the number of times that pattern n was used to update w during training. The corresponding predictions made by the trained Perceptron are therefore given by

$y(x) = sign(w^T \phi(x))$

$= sign(\sum_{n=1}^{N} \alpha_n t_n \phi(x_n)^T \phi(x))$

$= sign(\sum_{n=1}^{N} \alpha_n t_n k(x_n, x))$

Obviously, the prediction function of Perceptron has been expressed in terms of kernel function.

Then the learning algorithm of Perceptron is:

$\alpha_n \to \alpha_n + 1$

for the patterns which satisfy:

$t_n(\sum_{m=1}^{N} \alpha_m t_m \phi(x_m)^T \phi(x_n)) \geq 0$

And $\alpha_n \geq 0$so:

$t_n(\sum_{m=1}^{N} \phi(x_m)^T \phi(x_n)) \geq 0$

$t_n(\sum_{m=1}^{N} k(x_m, x_n)) \geq 0$

Bishop 7.3

Assume that a data set of two data point, x_1 and x_2, and they are in different class

+1 and -1 correspondingly then:

$w^T x_1 + b = +1$

$w^T x_2 + b = -1$

then we introduce lagrange multipliers λ and η

$argmin_{w,b} \frac{1}{2}||w||^2 + \lambda(w^T x_1 + b - 1) + \eta(w^T x_2 + b + 1)$

Then taking derivative w.r.t w and b and set to 0:

$0 = w + \lambda x_1 + \eta x_2$

$0 = \lambda + \eta$

Combining these two equation:

$w = \lambda(x_1 + x_2)$

Also combining the constraints of two data points, we get:

$2b = -\lambda^T(x_1 + x_2)$

Then combining with two equations above them, we get:

$b = -\frac{\lambda}{2}(x_1 + x_2)^T(x_1 + x_2)$

$= -\frac{\lambda}{2}(x_1^T x_1 + x_2^T x_2)$

Bishop 7.4

The margin is:

$\rho = \frac{1}{||w||}$

In order to get the maximum margin solution, we would vanish the second term

of:
$L(w,b,a) = \frac{1}{2}||w||^2 - \sum_{n=1}^{N} a_n t_n (w^T \phi(x_n) + b) - 1$
Then we get
$L(w,b,a) = \frac{1}{2}||w||^2$
and together with
$w = \sum_{n=1}^{N} a_n t_n \phi(x_n)$
we can write
$L(a) = \sum_{n=1}^{N} a_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} a_n a_m t_n t_m k(x_n, x_m)$
$\frac{1}{2}||w||^2 = \sum_{n}^{N} a_n - \frac{1}{2}||w||^2$
$||w||^2 = \sum_{n}^{N} a_n$ and together with
$\rho = \frac{1}{||w||}$
can get
$\frac{1}{\rho^2} = \sum_{n}^{N} a_n$