Bishop 6.2

4.55 change in the weight vector $w$ is

$w^{(T+1)} = w^{(T)} - \eta \nabla E_p(w) = w^{(T)} + \eta \phi_n t_n$

$\eta$ is the learning rate and $T$ is an integer that indexes the steps of the algorithm

The perceptron algorithm: two-class model feature vector $\phi(x)$ $(\phi_0(x)=1)$ generate a Linear model $y(x) = f(w^T\phi(x))$

nonlinear activation function $f(\,)$ is $f(a) = \begin{cases} +1 & a \geq 0 \\ -1 & a < 0 \end{cases}$

$w = \sum_{n=1}^{N} \alpha_n t_n \phi(x_n)$ $\alpha_n$ is the integer specifying the number of times that pattern $n$ was used to update $w$.

s.t. generate model $y(x) = f(w^T\phi(x)) = f(\sum_{n=1}^{N} \alpha_n t_n \phi(x_n)^T \phi(x)) = f(\sum_{n=1}^{N} \alpha_n t_n k(x_n, x))$ $f(\,)$ is $f(a)\begin{cases} +1 & a \geq 0 \\ -1 & a < 0 \end{cases}$

patterns which satisfy $t_n(w^T\phi(x_n)) \geq 0$. $\alpha_n \geq 0$ s.t. $t_n(\sum_{m=1}^{N} k(x_m, x_n)) \geq 0$

s.t. the learning algorithm depends only on the elements of the Gram Matrix.

Bishop 7.3

suppose we have two data points from $\{-1, 1\}$. which is $x_1 \in C_+ (t_1 = +1)$ and $x_2 \in C_- (t_2 = -1)$

we need to solve $\arg\min_{w,b} \frac{1}{2}\|w\|^2$ subject to the constraints $w^T x_1 + b = +1$

$w^T x_2 + b = -1$

in order to solve it we introduce Lagrange multipliers $\lambda$ and $\eta$.

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

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

$0 = \lambda + \eta$

$w = \lambda(x_1 - x_2)$

$b = -\frac{1}{2} w^T (x_1 + x_2)$

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

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

$\lambda$ still not decide s.t the magnitude of $w, b$ are not determined.

Bishop 7.4

Gives the dual representation of maximum margin problem

$\max \tilde{L}(a) = \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m t_n t_m k(x_n, x_m)$ kernel function $k(x, x') = \phi(x)^T \phi(x')$

subject to $a_n \geq 0$. $n = 1, 2 \cdots N$

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

$\rho = \frac{1}{\|w\|}$ $\frac{1}{\rho^2} = \|w\|^2$

$L(w, b, a) = \frac{1}{2}\|w\|^2$ for the maximum margin solution.

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

$\frac{1}{2}\|w\|^2 = \sum_{n=1}^{N} a_n - \frac{1}{2}\|w\|^2$

s.t. $\frac{1}{\rho^2} = \|w\|^2 = \sum_{n=1}^{N} a_n$