(1)

Problem set 4

PAUL NGOUCHET

CS 542

1. Written problems 6.2, 7.3, 7.4

(a) bishop 6.2.

$$(201)\quad w = \sum_{n=1}^{N} \alpha_n t_n \phi(x_n)$$

$$\begin{aligned} y(x) &= \operatorname{sign}(w^T \phi(x)) \\ &= \operatorname{sign}\left(\sum_{n=1}^{N} \alpha_n t_n \phi(x_n)^T \phi(x)\right) \\ &= \operatorname{sign}\left(\sum_{n=1}^{N} \alpha_n t_n k(x_n, x)\right) \end{aligned}$$

Thus the predictive function of the perceptron has been expressed purely in terms of the kernel function.

The learning algorithm of the perceptron can similarly be written as $\alpha_n \to \alpha_n + 1$

(2)

$$t_n \left(w^T \phi(x_n) \right) \geq 0$$

using $\alpha_n \geq 0$

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

so the learning algorithm depends only on the elements of the Gram matrix.

7.3)

given two datapoints $x_1 \in C_+ (t_1 = +1)$ and $x_2 \in C_- (t_2 = -1)$

$$w^T x_1 + b = +1$$

$$w^T x_2 + b = -1$$

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

Taking the derivative.

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

$$0 = \lambda + \eta$$

$$(220) \quad W = \lambda (X_1 - X_2) \qquad (3)$$

For b, we first rearrange and sum (216) and (217) to obtain

$$2b = -W^T (X_1 + X_2)$$

using (220)

$$b = -\frac{\lambda}{2} (X_1 - X_2)^T (X_1 + X_2)$$

$$\boxed{b = -\frac{\lambda}{2} (X_1^T X_1 - X_2^T X_2)}$$

λ is undetermined which reflects the inherent indeterminacy in the magnitude of W and b.

7.4)

From Figure 4.1 and (7.4)

we have $\Rightarrow \rho = \frac{1}{\|W\|}$ and so $\frac{1}{\rho^2} = \|W\|^2$

From 7.16 and (7.7)

$$L(w, b, a) = \frac{1}{2} \|W\|^2$$

(4)

using together (7.8) and (7.10)

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