Xiao Zhou zhouxiao@bu.edu

CS 542 Problem Set 4

Discussed with classmates

4/12/2017

1. Written Problems
   a. Bishop 6.2

      If we set w = 0, then the last w can be represented as

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

      Where α_n specify the number of times pattern n is used in training

      The prediction is given by

      $$y(x) = sign\big(w^T \phi(x)\big) = sign\left(\sum_{n=1}^{N} \alpha_n t_n \phi(x_n)^T \phi(x)\right)$$

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

      Where $k(x_n, x)$ is the kernel function, and the learning algorithm would be

      $\alpha_n = \alpha_n + 1$

      For misclassified patterns, we can have

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

      Then

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

      Which can be considered as

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

      Thus the learning algorithm only depend on the elements of the gram matrix.

   b. Bishop 7.3

      Suppose a data set has only 2 points, $x_1\ and\ x_2$, each belong to one class, then maximum margin hyperplane can be find by

      $$argmin_{w,b} \frac{1}{2} \left|\left|w\right|\right|^2$$

      Where

      $w^T x_1 + b = 1, w^T x_2 + b = -1$

Using Lagrange multipliers,

$$argmin_{w,b}\left(\frac{1}{2}\left||w|\right|^2+\lambda(w^Tx_1+b-1)+\eta(w^Tx_2+b+1)\right)$$

$$\frac{d}{dw}argmin_{w,b}\left(\frac{1}{2}\left||w|\right|^2+\lambda(w^Tx_1+b-1)+\eta(w^Tx_2+b+1)\right)=0$$

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

$$\frac{d}{db}argmin_{w,b}\left(\frac{1}{2}\left||w|\right|^2+\lambda(w^Tx_1+b-1)+\eta(w^Tx_2+b+1)\right)=0$$

$$\lambda+\eta=0$$

$$w=\lambda(x_1-x_2)$$

$$w^Tx_1+b+\ w^Tx_2+b=0$$

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

$$b=-\frac{\lambda}{2}(x_1-x_2)^T(x_1+x_2)=-\frac{\lambda}{2}(x_1^Tx_1-x_2^Tx_2)$$

So we can have

$$w=\lambda(x_1-x_2)$$

and

$$b=-\frac{\lambda}{2}(x_1^Tx_1-x_2^Tx_2)$$

c. Bishop 7.4

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

Since $a_n(t_ny(x_n)-1)=0$,

For max margin solution,

$$L(w,b,a)=\frac{1}{2}\left||w|\right|^2-\sum_{n=1}^{N}a_n(t_n(w^T\phi(x_n)+b)-1)$$

$$=\frac{1}{2}\left||w|\right|^2-\sum_{n=1}^{N}a_n(t_ny(x_n)-1)=\frac{1}{2}\left||w|\right|^2-0$$

Thus from 7.10 we can have

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

Consequently,

$$\frac{1}{\rho^2}=\sum_{n=1}^{N}a_n$$

2. Programming

   Codes and reports are in ‘.m’ files and ‘report.docx/pdf’ file.

   The code takes a while to complete, time varies based on the hardware.