

NNSP-1

Homework #4

Instructor: Prof., Shayan Garani Srinivasa

TA: Prayag Gowgi

Solution 1

Hyperplane is given by

$$w^T x + b = 0 \quad (1)$$

with the condition

$$\min_{i=1 \dots n} |w^T x_i + b| = 1 \quad (2)$$

Assuming w and b are optimum values, the discriminant function $g(x)$ gives the distance from the hyperplane

$$g(x) = w^T x + b \quad (3)$$

An alternate way to express x as a projection onto the hyperplane is given by

$$x = x_p + r \frac{w}{\|w\|} \quad (4)$$

where, x_p is the normal projection of x on to the hyperplane, r is the distance, positive if above and negative if below. Using this in equation (3) we get

$$r = \frac{g(x)}{\|w\|}. \quad (5)$$

Given the minimum condition it can be interpreted as

$$w^T x_i + b \geq 1, \text{ or } w^T x_i + b < -1 \text{ (for all } i = 1 \dots N) \quad (6)$$

$$r = \frac{1}{\|w_0\|} \text{ or } \frac{-1}{\|w_0\|} \quad (7)$$

Therefore,

$$\text{Minimum margin} = \frac{2}{\|w_0\|}. \quad (8)$$

Solution 2

$$K = \{k_{ij}\}_{i,j=1}^N, \quad k_{ij} = k(x_i, x_j). \quad (9)$$

K is positive and all elements k_{ij} are positive.

$$K = Q \Lambda Q^T \quad (10)$$

where, Q is eigen vector matrix, Λ is a diagonal matrix with eigenvalues and input space is of dimension m .

From Mercer's theorem,

$$K(x_i, x_j) = \sum_{k=1}^{\infty} \lambda_k Q_k(x_i) Q_k(x_j) \quad (11)$$

where λ_i are the eigen values, and Q_k are the eigen vectors.

Consider a feature map

$$\Phi(x_i) = \left[\sqrt{\lambda_1} \Phi_1^i, \sqrt{\lambda_2} \Phi_2^i, \sqrt{\lambda_t} \Phi_t^i \dots \right]_n \quad (12)$$

and inner product between $\Phi(x_i)$ and $\Phi(x_j)$ is given by

$$\begin{aligned} \langle \Phi(x_i) \Phi(x_j) \rangle &= \sum_{t=1}^n \lambda_t \Phi_t^i \Phi_t^j = K_{ij} \\ &= (Q \Lambda Q^T)_{ij} \end{aligned}$$

However, we know that K_{ij} are positive for all i, j .

Solution 3

(a) (a)

$$K(x, x_j) = (x^T x_j + 1)^P \quad (13)$$

$$\begin{aligned} K(Qx, Qx_j) &= ((Qx)^T (Qx_j) + 1)^P \\ &= (Q^T Q x^T x_j + 1)^P \end{aligned}$$

Given $Q^T = Q^{-1}$, $Q^T Q = 1$,

$$K(Qx, Qx_j) = (x^T x_j + 1)^P = K(x, x_j). \quad (14)$$

(b)

$$K(x, x_j) = \exp\left(-\frac{1}{2\sigma^2} \|x - x_i\|^2\right) \quad (15)$$

$$\begin{aligned} K(Qx, Qx_j) &= \exp\left(-\frac{1}{2\sigma^2} \|Qx - Qx_i\|^2\right) \\ &= \exp\left(-\frac{1}{2\sigma^2} \|Q(x - x_i)\|^2\right) \end{aligned}$$

Taking Q as a constant outside the norm,

$$\begin{aligned} K(Qx, Qx_j) &= \exp\left(-\frac{1}{2\sigma^2} \left(\sqrt{Q^T Q}\right)^2 \|x - x_i\|^2\right) \\ &= \exp\left(-\frac{1}{2\sigma^2} \|x - x_i\|^2\right) = K(x, x_j) \end{aligned}$$

(c)

$$K(x, x_j) = \tanh(\beta x^T x_j + \beta_1) \quad (16)$$

$$\begin{aligned} K(Qx, Qx_j) &= \tanh(\beta(Qx)^T (Qx_j) + \beta_1) \\ &= \tanh(\beta x^T x_j + \beta_1) \text{ since } Q^T Q = 1. \\ &= K(x, x_j) \end{aligned}$$

(b) Yes, invariance in the input space is maintained in the feature space.

Solution 4

We generated the data as shown in Figure (1) of HW4. We used package `e1071` in R to perform this experiment. The kernel is polynomial type. We were able to classify the data into two separate classes. Code is attached.

Solution 5

- a The vertical separation $d = -6.5$ and the parameter value C is the regularization parameter. We would like to find an optimum C that minimizes the classification error. We have used package `e1071` in R to perform this experiment. The kernel used is radial type. From the table, we see that $C = 0.7$ results in minimum error.

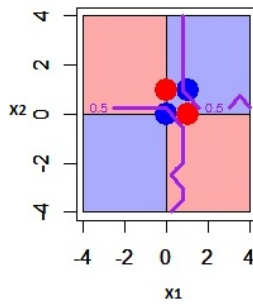
C	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
No. of misclassification	31	12	11	8	7	4	3	3	3	2
Error	0.0517	0.0200	0.0183	0.0133	0.0117	0.0067	0.0050	0.0050	0.0050	0.0033

- b The vertical separation $d = -6.75$ and the parameter value C is the regularization parameter. We would like to find an optimum C that minimizes the classification error. We have used package `e1071` in R to perform this experiment. The kernel used is radial type. From the table, we see that $C = 10$ results in minimum error.

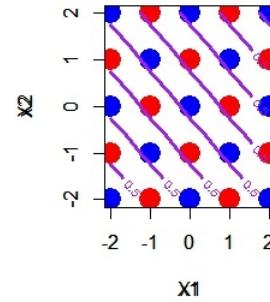
C	0.1	0.5	1	5	10	20	30	40	50
No. of misclassification	46	11	6	4	2	2	2	2	2
Error	0.0767	0.0200	0.0100	0.0067	0.0033	0.0033	0.0033	0.0033	0.0033

Solution 6

We have used package `e1071` in R to perform this experiment. The kernel used is radial type. We performed two experiments a) simple XOR b) Tiled XOR. Following are the decision boundaries for the two experiments.



(a) Decision boundary for simple XOR



(b) Decision boundary for tiled XOR

FIGURE 1 – Decision boundary for XOR problem

A Code for Q4

```
xors<-read.table("xor_data_3d.txt",header=T)
xors$label<-as.factor(xors$label-1)
xors.svm<-svm(label~.,xors,type='C-classification', kernel='polynomial')
par(mfrow=c(1,2))
par(new=T)
px<-seq(-2,2,1)
py<-seq(-2,2,1)
pz <- seq(-2,2,1)
pgrid<-expand.grid(px,py,pz)
colnames(pgrid) <- c("x", "y", "z")
contour(px,py, array(predict(xors.svm,newdata=xors[,1:3]),dim = c(length(px),length(py))),
),levels = 0.5,drawlabels = T,col='purple',lwd=3,xlim=c(-2,2),ylim=c(-2,2))
```