R05631027_楊皓文_機器學習_作業#7

Problem 1 (Support vector machine)

r05631027 楊皓文 ML#7

Date

1. (a)

$$\phi\left(\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right)=\begin{bmatrix}\alpha\\ \beta\\ \alpha^2\end{bmatrix}$$

$$\phi(x_1)=\phi\left(\begin{bmatrix}-1\\ 1\end{bmatrix}\right)=\begin{bmatrix}-1\\ 1\\ 1\end{bmatrix}$$

$$\phi(x_2)=\phi\left(\begin{bmatrix}0\\ 1\end{bmatrix}\right)=\begin{bmatrix}0\\ 1\\ 0\end{bmatrix}$$

$$\phi(x_3)=\phi\left(\begin{bmatrix}1\\ 1\end{bmatrix}\right)=\begin{bmatrix}1\\ 1\\ 1\end{bmatrix}$$

$$\phi(x_4)=\phi\left(\begin{bmatrix}0\\ 0\end{bmatrix}\right)=\begin{bmatrix}0\\ 0\\ 0\end{bmatrix}$$

(b)

$$K(x_1,x_1)=\begin{bmatrix}-1&1&1\end{bmatrix}\begin{bmatrix}-1\\ 1\\ 1\end{bmatrix}=3$$

$$K(x_1,x_2)=\begin{bmatrix}-1&1&1\end{bmatrix}\begin{bmatrix}0\\ 1\\ 0\end{bmatrix}=1$$

$$K(x_1,x_3)=\begin{bmatrix}-1&1&1\end{bmatrix}\begin{bmatrix}1\\ 1\\ 1\end{bmatrix}=1$$

$$K(x_1,x_4)=\begin{bmatrix}-1&1&1\end{bmatrix}\begin{bmatrix}0\\ 0\\ 0\end{bmatrix}=0$$

$$K(x_2,x_2)=\begin{bmatrix}0&1&0\end{bmatrix}\begin{bmatrix}0\\ 1\\ 0\end{bmatrix}=1$$

$$K(x_2,x_3)=\begin{bmatrix}0&1&0\end{bmatrix}\begin{bmatrix}1\\ 1\\ 1\end{bmatrix}=1$$

$$K(x_2,x_4)=\begin{bmatrix}0&1&0\end{bmatrix}\begin{bmatrix}0\\ 0\\ 0\end{bmatrix}=0$$

$$K(x_3,x_3)=\begin{bmatrix}1&1&1\end{bmatrix}\begin{bmatrix}1\\ 1\\ 1\end{bmatrix}=3$$

$$K(x_3,x_4)=\begin{bmatrix}1&1&1\end{bmatrix}\begin{bmatrix}0\\ 0\\ 0\end{bmatrix}=0$$

$$K(x_4,x_4)=\begin{bmatrix}0&0&0\end{bmatrix}\begin{bmatrix}0\\ 0\\ 0\end{bmatrix}=0$$

(c)

$$L(w,b,\lambda_i)=\frac{1}{2}\|w\|^2-\sum_{i=1}^{4}\lambda_i\left[y_i(w^T\phi_i+b)-1\right]$$

$$=\frac{1}{2}\|w\|^2-\sum_{i=1}^{4}\lambda_i y_i(w^T\phi_i+b)+\sum_{i=1}^{4}\lambda_i$$

$$w=\sum_{i=1}^{4}\lambda_i y_i\phi_i$$

$$b=1-w^T\phi_i$$

Chrys culture

(d)
$$\begin{cases} \dfrac{\partial L}{\partial w} = w - \sum_{i=1}^{4} \lambda_i y_i \phi_i = 0 \\ \dfrac{\partial L}{\partial b} = \sum_{i=1}^{4} \lambda_i y_i = 0 \end{cases}$$

$$\Rightarrow L_d = \sum_{i=1}^{4} \lambda_i - \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \lambda_i \lambda_j y_i y_j \phi_i^T \phi_j$$

$$\text{, subject to } \begin{cases} \sum_{i=1}^{4} \lambda_i y_i = 0 \\ \lambda_i \geq 0 , \forall i \end{cases}$$

$$L_n(\lambda_i, \alpha) = f(\lambda_i) - \alpha g(\lambda_i)$$
$$= \sum_{i=1}^{4} \lambda_i - \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \lambda_i \lambda_j y_i y_j \phi_i^T \phi_j - \alpha \sum_{i=1}^{4} \lambda_i y_i$$
$$= \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 - \frac{1}{2}\Big(\lambda_1 \lambda_1 y_1 y_1 \phi_1^T \phi_1 + \lambda_1 \lambda_2 y_1 y_2 \phi_1^T \phi_2 + \lambda_1 \lambda_3 y_1 y_3 \phi_1^T \phi_3$$
$$+ \cancel{\lambda_1 \lambda_4 y_1 y_4 \phi_1^T \phi_4} + \lambda_2 \lambda_2 y_2 y_2 \phi_2^T \phi_2 + \lambda_2 \lambda_3 y_2 y_3 \phi_2^T \phi_3$$
$$+ \cancel{\lambda_2 \lambda_4 y_2 y_4 \phi_2^T \phi_4} + \lambda_3 \lambda_3 y_3 y_3 \phi_3^T \phi_3 + \cancel{\lambda_3 \lambda_4 y_3 y_4 \phi_3^T \phi_4}$$
$$+ \cancel{\lambda_4 \lambda_4 y_4 y_4 \phi_4^T \phi_4}\Big) - \alpha(\lambda_1 - \lambda_2 + \lambda_3 + \lambda_4)$$
$$= \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 - \frac{3}{2}\lambda_1^2 + \frac{1}{2}\lambda_1\lambda_2 - \frac{1}{2}\lambda_1\lambda_3 - \frac{1}{2}\lambda_2^2 + \frac{1}{2}\lambda_2\lambda_3 - \frac{3}{2}\lambda_3^2$$
$$- \alpha(\lambda_1 - \lambda_2 + \lambda_3 + \lambda_4)$$

$$\Rightarrow \begin{cases} \dfrac{\partial L_n}{\partial \lambda_1} = 1 - 3\lambda_1 + \frac{1}{2}\lambda_2 - \frac{1}{2}\lambda_3 - \alpha = 0 \\ \dfrac{\partial L_n}{\partial \lambda_2} = 1 + \frac{1}{2}\lambda_1 - \lambda_2 + \frac{1}{2}\lambda_3 + \alpha = 0 \\ \dfrac{\partial L_n}{\partial \lambda_3} = 1 - \frac{1}{2}\lambda_1 + \frac{1}{2}\lambda_2 - 3\lambda_3 - \alpha = 0 \\ \dfrac{\partial L_n}{\partial \lambda_4} = 1 - \alpha = 0 \\ \dfrac{\partial L_n}{\partial \alpha} = \lambda_1 - \lambda_2 + \lambda_3 + \lambda_4 = 0 \end{cases}$$

$$\Rightarrow \alpha = 1 , \lambda_1 = \frac{1}{3} , \lambda_2 = \frac{7}{3} , \lambda_3 = \frac{1}{3} , \lambda_4 = \frac{5}{3}$$

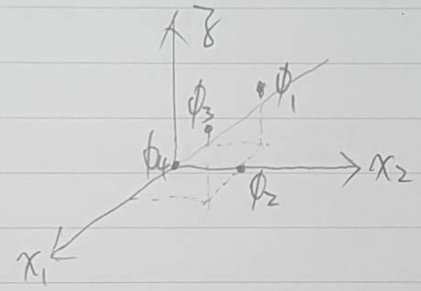
No.
Date

(e) $w = \sum_{i=1}^{4} \lambda_i y_i \phi_i = \frac{1}{3} \cdot 1 \cdot \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} + \frac{7}{3} \cdot (-1) \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \frac{1}{3} \cdot 1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{5}{3} \cdot 1 \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$= \begin{bmatrix} 0 \\ -\frac{5}{3} \\ \frac{2}{3} \end{bmatrix}$

$b = 1 - w^T \phi_i = 1 - \begin{bmatrix} 0 & -\frac{5}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = 2$

(f) $f(x) = \mathrm{sgn}(w^T \phi + b) = \mathrm{sgn}(\begin{bmatrix} 0 & -\frac{5}{3} & \frac{2}{3} \end{bmatrix} \phi + 2)$



Problem 2 (Support vector machine)

(a) Develop an SVM classifier to differentiate the images of digit 0 to the images of digit 1. Use softmargin SVM classifier and an RBF kernel.

cost \ gamma	2^{-14}	2^{-13}	2^{-12}	2^{-11}	2^{-10}	2^{-9}	2^{-8}	2^{-7}	2^{-6}
2^{-5}	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5
2^{-4}	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5
2^{-3}	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5
2^{-2}	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5
2^{-1}	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	100
1	99.5	99.5	99.5	99.5	99.5	99.5	99.5	100	100
2	99.5	99.5	99.5	99.5	99.5	99.5	100	100	100
2^{2}	99.5	99.5	99.5	99.5	99.5	100	100	100	100
2^{3}	99.5	99.5	99.5	99.5	100	100	100	100	100

10 fold-validation

cost \ gamma	2^{-14}	2^{-13}	2^{-12}	2^{-11}	2^{-10}	2^{-9}	2^{-8}	2^{-7}	2^{-6}
2^{-5}	92.9	92.9	92.9	92.9	92.9	92.9	92.9	92.9	92.9
2^{-4}	92.9	92.9	92.9	92.9	92.9	92.9	92.9	92.9	92.9
2^{-3}	92.9	92.9	92.9	92.9	92.9	92.9	92.9	92.9	92.9
2^{-2}	92.9	92.9	92.9	92.9	92.9	92.9	92.9	92.9	92.9
2^{-1}	92.9	92.9	92.9	92.9	92.9	92.9	92.9	92.9	97
1	93	93	93	93	93	93	93	97	98.9
2	93	93	93	93	93	93	97	98.9	99.6
2^{2}	93	93	93	93	93	97	98.9	99.6	99.7
2^{3}	93.1	93.1	93.1	93.1	97	98.9	99.6	99.7	99.7

Discussion:

我們由 test data 丟入 model 可以發現，雖然差別不大(99.5~100%)，但大致趨勢是 cost 和 gamma 越大，其準確率也越高。

而從 training data 的 10 fold cross-validation，則可以看到較大的差別(92.9%~99.7%)，同樣的也呈現，當 cost 和 gamma 越大，其準確率也越高。

(b) Try linear and polynomial kernels with various kernel parameters. Compare and report the performance of the kernels on the test data.

Kernel: linear(-t 0)

cost \ gamma	2^{-14}	2^{-13}	2^{-12}	2^{-11}	2^{-10}	2^{-9}	2^{-8}	2^{-7}	2^{-6}

2^{-5}	100	100	100	100	100	100	100	100	100
2^{-4}	100	100	100	100	100	100	100	100	100
2^{-3}	100	100	100	100	100	100	100	100	100
2^{-2}	100	100	100	100	100	100	100	100	100
2^{-1}	100	100	100	100	100	100	100	100	100
1	100	100	100	100	100	100	100	100	100
2	100	100	100	100	100	100	100	100	100
2^{2}	100	100	100	100	100	100	100	100	100
2^{3}	100	100	100	100	100	100	100	100	100

Kernel: linear and cross-validation(-t 0 –v 10)

cost \ gamma	2^{-14}	2^{-13}	2^{-12}	2^{-11}	2^{-10}	2^{-9}	2^{-8}	2^{-7}	2^{-6}
2^{-5}	98.9	99.6	99.7	99.7	99.6	99.6	99.8	99.8	99.8
2^{-4}	98.9	99.6	99.7	99.7	99.6	99.6	99.8	99.8	99.8
2^{-3}	98.9	99.6	99.7	99.7	99.6	99.6	99.8	99.8	99.8
2^{-2}	98.9	99.6	99.7	99.7	99.6	99.6	99.8	99.8	99.8
2^{-1}	98.9	99.6	99.7	99.7	99.6	99.6	99.8	99.8	99.8
1	98.9	99.6	99.7	99.7	99.6	99.6	99.8	99.8	99.8
2	98.9	99.6	99.7	99.7	99.6	99.6	99.8	99.8	99.8
2^{2}	98.9	99.6	99.7	99.7	99.6	99.6	99.8	99.8	99.8
2^{3}	98.9	99.6	99.7	99.7	99.6	99.6	99.8	99.8	99.8

Kernel: polynomial(-t 1)

cost \ gamma	2^{-14}	2^{-13}	2^{-12}	2^{-11}	2^{-10}	2^{-9}	2^{-8}	2^{-7}	2^{-6}
2^{-5}	93	93	93	93	93	93	93	93	93
2^{-4}	93	93	93	93	93	93	93	93	93
2^{-3}	93	93	93	93	93	93	93	93	93
2^{-2}	93	93	93	93	93	93	93	93	93
2^{-1}	93	93	93	93	93	93	93	93	93
1	93	93	93	93	93	93	93	93	93
2	93	93	93	93	93	93	93	93	93

2^2	93	93	93	93	93	93	93	93	93
2^3	93	93	93	93	93	93	93	93	93

Kernel: polynomial and cross-validation(-t 1 –v 10)

cost \ gamma	2^{-14}	2^{-13}	2^{-12}	2^{-11}	2^{-10}	2^{-9}	2^{-8}	2^{-7}	2^{-6}
2^{-5}	50	50	50	50	50	50	50	50	50
2^{-4}	50	50	50	50	50	50	50	50	50
2^{-3}	50	50	50	50	50	50	50	50	50
2^{-2}	50	50	50	50	50	50	50	50	50
2^{-1}	50	50	50	50	50	50	50	50	50
1	50	50	50	50	50	50	50	50	50
2	50	50	50	50	50	50	50	50	50
2^2	50	50	50	50	50	50	50	50	50
2^3	50	50	50	50	50	50	50	50	50

Discussion:

由上圖分別做 linear 和 polynomial kernel 可以發現，linear 的 kernel 的辨識準確率較高，甚至大於(a)小題使用 RBF 的準確率。上網查詢資料後，發現可能因為當 number of features 很大時(此題為 28x28=784)，由 RBF kernel 投影到更高維度並不會增加準確率，反而 linear 會有較好的表現。