10601 Machine Learning HW4
Hsueh Lin Huang
hsuehlih

1. Support Vector Machines

1.1 Feature Maps

1. No, because we can't use one line to seperate blue and red points. Therefore, it's not linear separable.

2. Yes, $x_2 - 2.5 \geq 0$, the corresponding margin is 1.5.

3. $K(x,z) = x\,z + x^2 z^2$

1.2 Constructing Kernels

1. $K(x,z) = c_1 k_1(x,z) + c_2 k_2(x,z)$ with $c_1, c_2 \geq 0$

$\phi(x) = (\sqrt{c_1}\,\phi_1(x), \sqrt{c_2}\,\phi_2(x))$

$\phi(x)\cdot\phi(z) = c_1 \underbrace{\phi_1(x)\cdot\phi_1(z)}_{k_1(x,z)} + c_2 \underbrace{\phi_2(x)\cdot\phi_2(z)}_{k_2(x,z)}$

2. $K(x,z) = k_1(x,z)\cdot k_2(x,z)$

$\phi(x) = (\phi_{1,i}(x)\ \phi_{2,j}(x))_{i\in\{1\ldots n\},\ j\in\{1\ldots m\}}$

$\phi(x)\cdot\phi(z) = \sum_{i,j} \phi_{1,i}(x)\,\phi_{2,j}(x)\,\phi_{1,i}(z)\,\phi_{2,j}(z)$

$= \sum_i \phi_{1,i}(x)\,\phi_{1,i}(z)\left(\sum_j \phi_{2,j}(x)\,\phi_{2,j}(z)\right)$

$= \sum_i \phi_{1,i}(x)\,\phi_{1,i}(z)\,k_2(x,z) = k_1(x,z)\,k_2(x,z)$

3.

$q(t)=\sum_{i=0}^{p} c_i t^i$, prove $K(x,z)=q(k_1(x,z))$

$$\sum_{i=0}^{p} c_i k_1(x,z)^i = c_0 k_1(x,z)^0 + c_1 k_1(x,z)^1 + \dots + c_p k_1(x,z)^p$$

Use $K(x,z)=k_1(x,z)k_2(x,z)$, we get:

$$\sum_{i=0}^{p} c_i k_1(x,z)^i = c_0 k^{(0)}(x,z) + c_1 k^{(1)}(x,z) + \dots + c_p k^{(p)}(x,z)$$

Then use $K(x,z)=c_1 k_1(x,z)+c_2 k_2(x,z)$, we get:

$$K(x,z)=q(k_1(x,z)) = \sum_{i=0}^{p} c_i k^{(i)}(x,z) = \sum_{i=0}^{p} c_i k_1(x,z)^i \ \#$$

4.

$K(x,z)=\exp(k_1(x,z))$

Since: $\exp(x)=\lim_{i\to\infty}\left(1+x+\dots+\frac{x^i}{i!}\right)$

From the fact that $k(x,z)=f(k_1(x,z))$,

$$K(x,z)=\lim_{i\to\infty} k_i(x,z)$$

5. $K(x,z)=x^T A z$

1.3 Support Vectors

1. $\sum_{i}^{N} \ell(w,x,y)=\max(0,1-y(w\cdot x))$

2. C controls the relative weighting between the twin goals of making the $\|w\|^2$ small (margin is large) and ensuring that most examples have functional margin ≥ 1

If $C \to 0$, r will be very large

$C \to \infty$, r will be very small

3. Classifier: $f(x)=\sum_{i}^{N}\alpha_i y_i \phi(x_i)\phi(x)+b$

$\Rightarrow f(x)=\sum_{i}^{N}\alpha_i y_i K(x_i,x)+b$

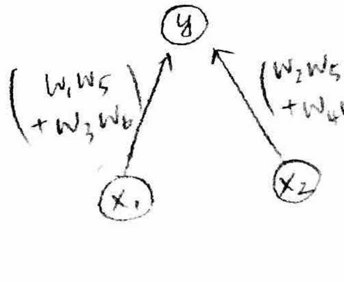
Learning $\max_{\alpha_i \geq 0} \sum \alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j \phi(x_i)^T\phi(x_j)$

$\Rightarrow \max_{\alpha_i\geq 0} \sum\alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j K(x_i,x_j)$

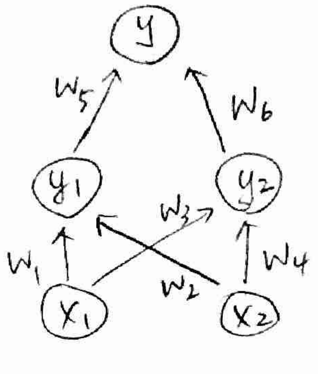
2. Feed-forward Neural Networks

1. Yes

$y=(w_1w_5+w_3w_6)x_1+(w_2w_5+w_4w_6)x_2+w_0$



2.



for (1,1)

$$(W_1+W_2)W_5+(W_3+W_4)W_6+W_0=1 \quad (1)$$

for (-1,1)

$$(-W_1+W_2)W_5+(-W_3+W_4)W_6+W_0=-1. \quad (2)$$

for (-1,-1)

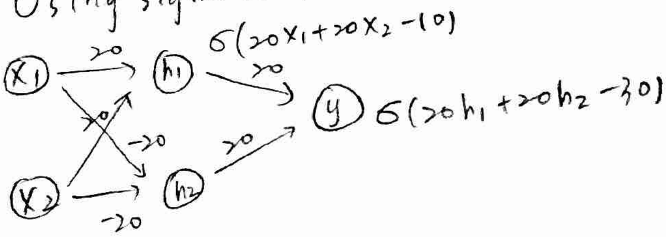
$$(-W_1-W_2)W_5+(-W_3-W_4)W_6+W_0=1 \quad . (3)$$

for (1,-1).

$$(W_1-W_2)W_5+(W_3-W_4)W_6+W_0=-1 \quad (4)$$

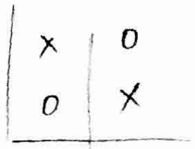
From (1),(3), we conclude that $W_0=1$
(2),(4) we conclude that $W_0=-1$] conflicts!

3.
Using sigmoid function, we can solve XOR problem.



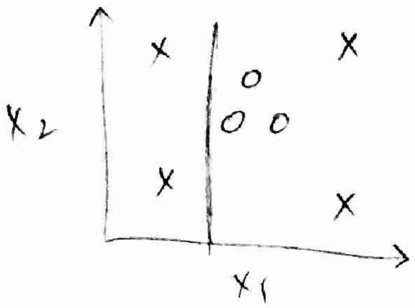
3. AdaBoost

1.



the minimum error rate is 0.5

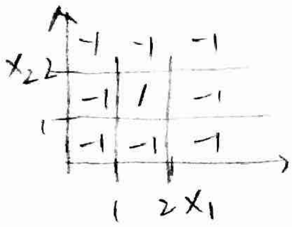
2、



$$h_1(x) = \begin{cases} 1 & \text{if } x_1 \leq 1 \\ -1 & \text{otherwise} \end{cases}$$

error rate $= \frac{2}{7}$

3、



① $x_1 < 1$, $x_2 > 2 \Rightarrow h_1 = -1$ $h_2 = 1$, $h_3 = 1$, $h_4 = -1$

② $2 \geq x_1 \geq 1$, $x_2 > 2 \Rightarrow h_1 = 1$, $h_2 = 1$, $h_3 = 1$, $h_4 = -1$

③ $x_1 > 2$, $x_2 > 2 \Rightarrow h_1 = 1$, $h_2 = -1$, $h_3 = 1$, $h_4 = -1$

④ $x_1 < 1$, $1 \leq x_2 \leq 2 \Rightarrow h_1 = -1$, $h_2 = 1$, $h_3 = 1$, $h_4 = 1$

⑤ $1 \leq x_1 \leq 2$, $1 \leq x_2 \leq 2 \Rightarrow h_1 = 1$, $h_2 = 1$, $h_3 = 1$, $h_4 = 1$

⑥ $x_1 > 2$, $1 \leq x_2 \leq 2 \Rightarrow h_1 = 1$, $h_2 = -1$, $h_3 = 1$, $h_4 = 1$

⑦ $x_1 < 1$, $x_2 < 1 \Rightarrow h_1 = -1$, $h_2 = 1$, $h_3 = -1$, $h_4 = 1$

⑧ $1 \leq x_1 \leq 2$, $x_2 < 1 \Rightarrow h_1 = 1$, $h_2 = 1$, $h_3 = -1$, $h_4 = 1$

⑨ $x_1 > 2$, $x_2 < 1 \Rightarrow h_1 = 1$, $h_2 = -1$, $h_3 = -1$, $h_4 = 1$

for ① $y_i f(x_i) = -1 \cdot \frac{1}{7}(0-3) = \frac{3}{7}$

② $y_i f(x_i) = -1 \cdot \frac{1}{7}(2-3) = \frac{1}{7}$

③ $y_i f(x_i) = -1 \cdot \frac{1}{7}(0-3) = \frac{3}{7}$

④ $y_i f(x_i) = -1 \cdot \frac{1}{7}(2-3) = \frac{1}{7}$

⑤ $y_i f(x_i) = 1 \cdot \frac{1}{7}(4-3) = \frac{1}{7}$

⑥ $y_i f(x_i) = -1 \cdot \frac{1}{7}(2-3) = \frac{1}{7}$

⑦ $y_i f(x_i) = -1 \cdot \frac{1}{7}(0-3) = \frac{3}{7}$

⑧ $y_i f(x_i) = -1 \cdot \frac{1}{7}(2-3) = \frac{1}{7}$

⑨ $y_i f(x_i) = -1 \cdot \frac{1}{7}(0-3) = \frac{3}{7}$

$\Rightarrow y_i f(x_i) \geq \frac{1}{7}$ #

3.2 Complexity and Overfitting

1. 4 params — dimension, threshold, direction, weight

2. $5 \times T$ params.

3. No. because the testing error doesn't increase.

3.3 Margin

1. $\text{Margin}_k(x) = \sum_{t: y = h_t(x)} a_t - \sum_{t: y \neq h_t(x)} a_t = y \cdot \sum a_t h_t(x) = y \cdot f_k(x)$

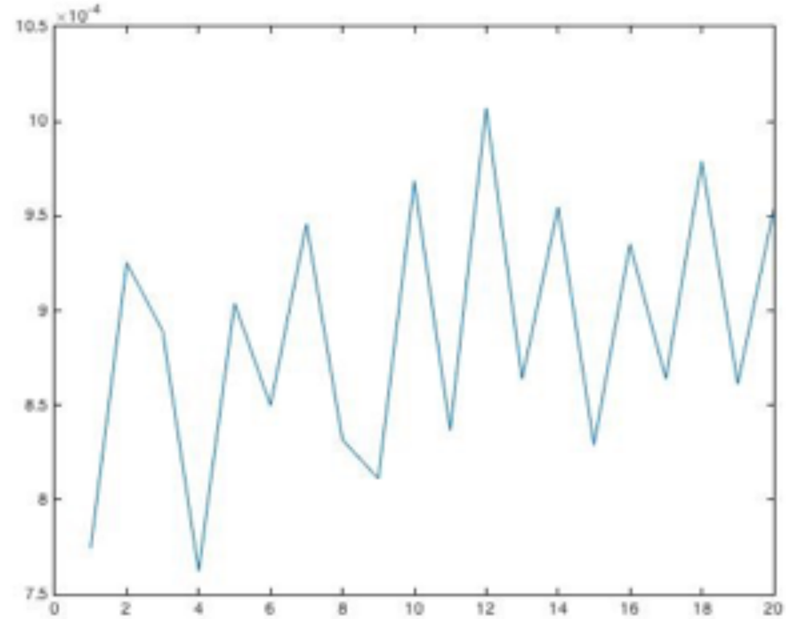
if $y = h_t(x) \Rightarrow y \cdot h_t(x) = 1$

$y \neq h_t(x) \Rightarrow y \cdot h_t(x) = -1$

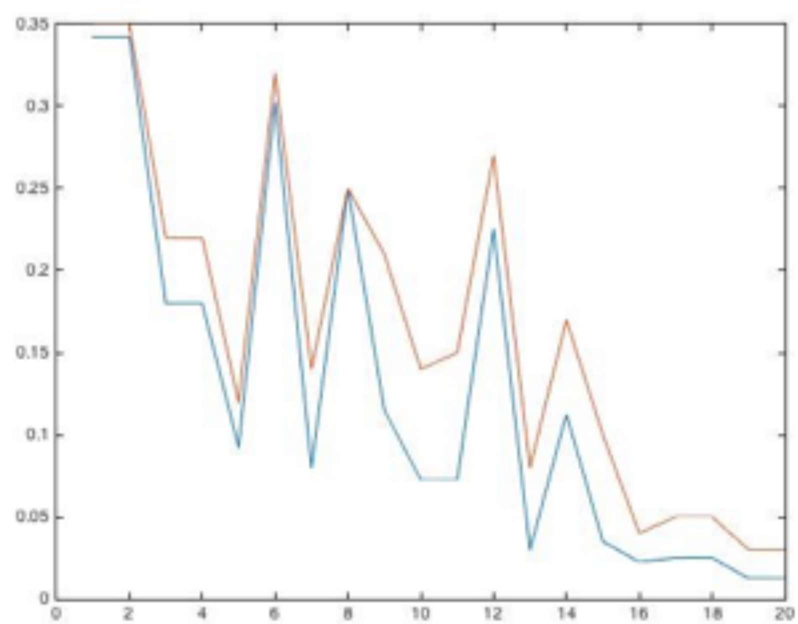
2.

3.5.1 Report Questions

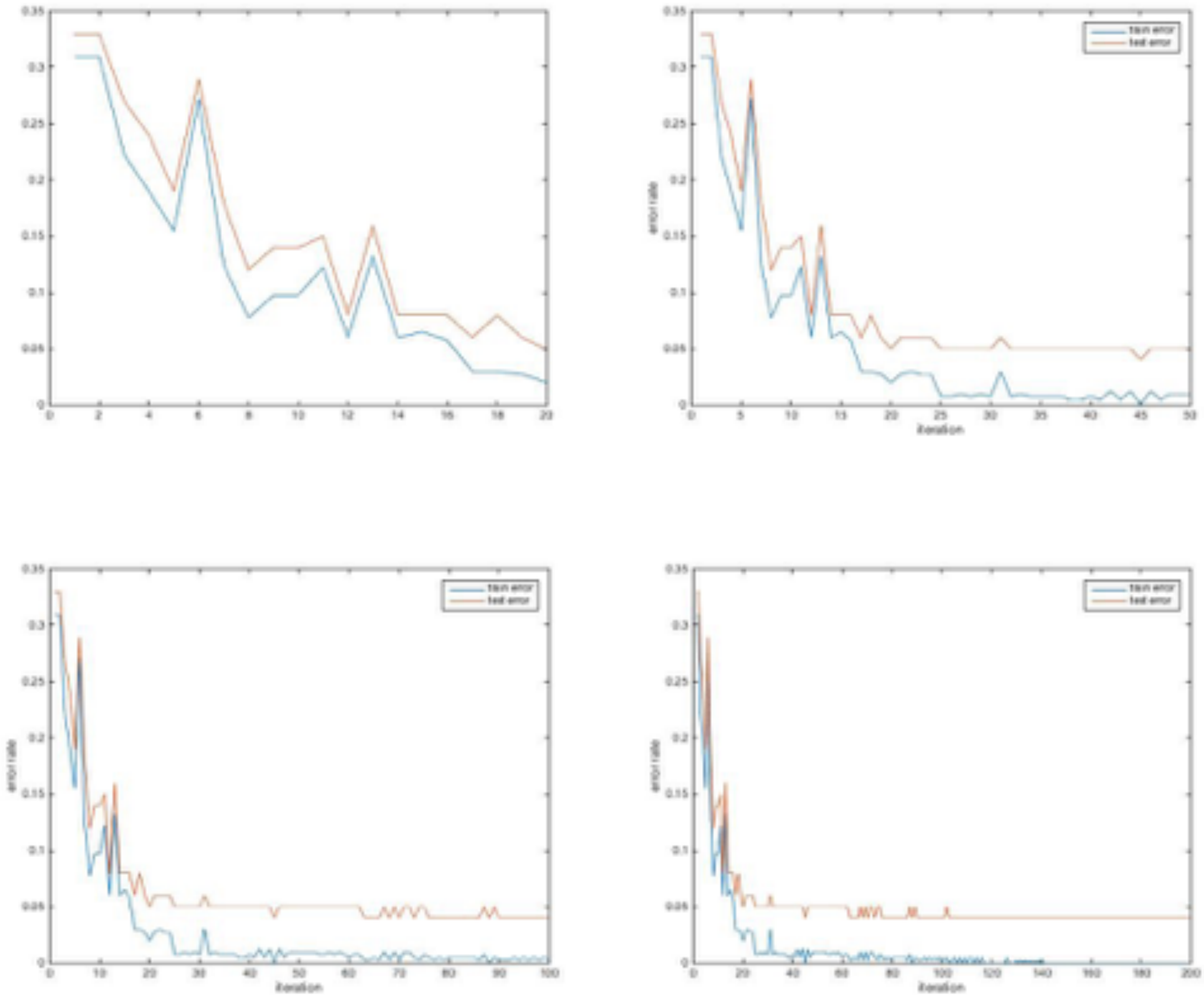
1. the highest weighted error rate: [h12, error rate = 0.0010]



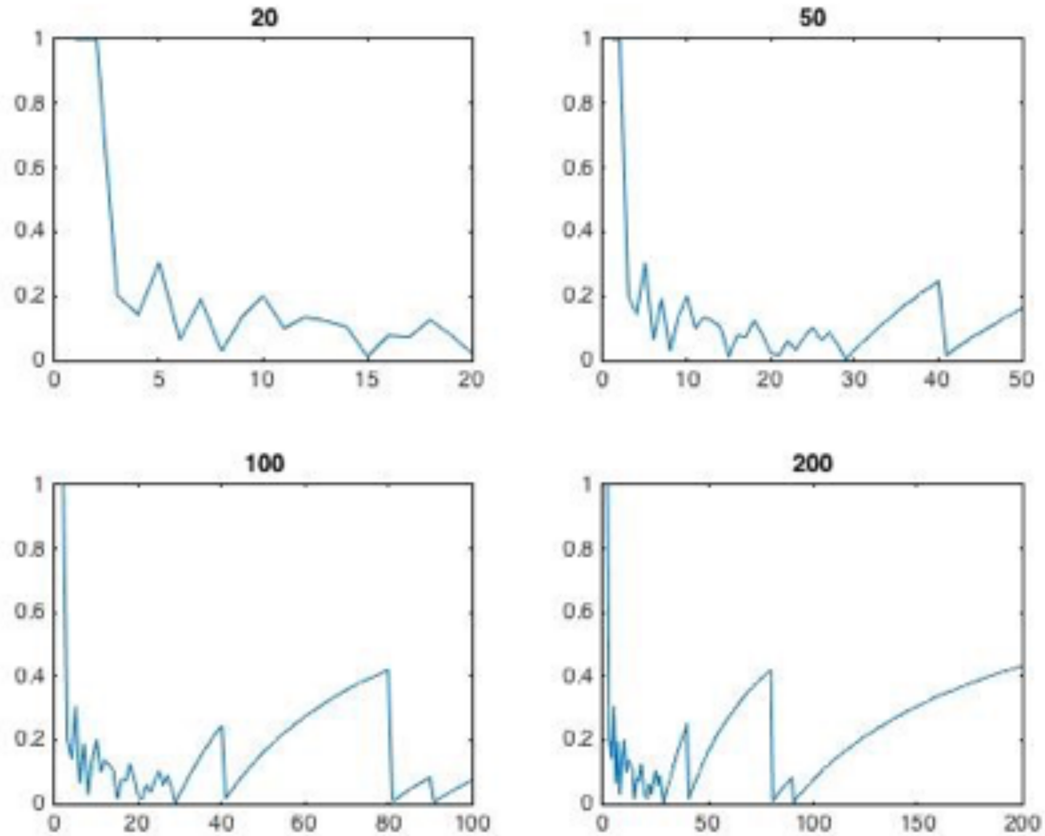
2. & 3.
T = 20



4.



5.



20
50
100
200