Date

Homework 4

2. (1) $\{-3,-2,-1,0,1,2,3\}$ all can be support vector. 7.

(2) $\{-1,0\}$ can be support vector. 2.

(3)(a) $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j) = \Phi(x_j)^T \Phi(x_i) = K(x_j, x_i)$, hence symmetric.

(b) Suppose $B = [\Phi(x_1), \Phi(x_2), \cdots, \Phi(x_n)]$, then $A = B^T B$

$$x^T A x = x^T B^T B x = (Bx)^T (Bx) = \|Bx\|^2 \geq 0$$, hence A semi-definite.

3. (1) '−' weight increase because G_1 make all be '+'.

(2) 3. Suppose the index of '−' is 1, '+' are 2, 3, 4, 5.

⟨0⟩ At the beginning, $\omega_i^{(1)} = \frac{1}{5}$.

⟨1⟩ $\varepsilon_1 = \frac{1}{5}$, $\alpha_1 = \log 4$, all samples of '+' don't change,

$\omega_1^{(2)}$ become $\frac{4}{5}$.

⟨2⟩ 2 '+' will be classify into '−'. $\varepsilon_2 = \frac{1}{4}$, $\alpha_2 = \log 3$.

Then $\omega_1^{(3)} = \frac{4}{5}$, $\omega_2^{(3)} = \frac{3}{5}$, $\omega_3^{(3)} = \frac{3}{5}$, $\omega_4^{(3)} = \frac{1}{5}$, $\omega_5^{(3)} = \frac{1}{5}$.

⟨3⟩ $\varepsilon_3 = \frac{1}{6}$, $\alpha_3 = \log 5$.

$$G(x) = \text{sign}(\log 4 \cdot G_1(x) + \log 3 \cdot G_2(x) + \log 5 \cdot G_3(x))$$

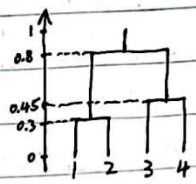
(3) No. After 2 iterations, there must have a '+' and a '−' in the same class. Hence can not reduce training error to 0.

4. (1)

$$\begin{array}{c|c} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 0 & 0.3 & 0.4 & 0.7 \\ 0.3 & 0 & 0.5 & 0.8 \\ 0.4 & 0.5 & 0 & 0.45 \\ 0.7 & 0.8 & 0.45 & 0 \end{pmatrix} \end{array}$$

$$\Downarrow$$

$$\begin{array}{c|c} & \begin{matrix} (1,2) & 3 & 4 \end{matrix} \\ \begin{matrix} (1,2) \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 0 & 0.5 & 0.8 \\ 0.5 & 0 & 0.45 \\ 0.8 & 0.45 & 0 \end{pmatrix} \end{array} \Rightarrow \begin{array}{c|c} & \begin{matrix} (1,2) & (3,4) \end{matrix} \\ \begin{matrix} (1,2) \\ (3,4) \end{matrix} & \begin{pmatrix} 0 & 0.8 \\ 0.8 & 0 \end{pmatrix} \end{array}$$



No.

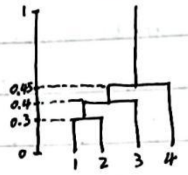
Date

(2)

$$\begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & & & & \\ 2 & & & & \\ 3 & & & & \\ 4 & & & & \end{array}\quad \begin{pmatrix} 0 & 0.3 & 0.4 & 0.7 \\ 0.3 & 0 & 0.5 & 0.8 \\ 0.4 & 0.5 & 0 & 0.45 \\ 0.7 & 0.8 & 0.45 & 0 \end{pmatrix}$$

$$\Downarrow$$

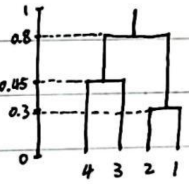
$$\begin{array}{c|c} & \begin{array}{ccc} (1,2) & 3 & 4 \end{array} \\ \hline \begin{array}{c} (1,2) \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & 0.4 & 0.7 \\ 0.4 & 0 & 0.45 \\ 0.7 & 0.45 & 0 \end{pmatrix} \end{array} \Rightarrow \begin{array}{c|c} & \begin{array}{cc} ((1,2),3) & 4 \end{array} \\ \hline \begin{array}{c} ((1,2),3) \\ 4 \end{array} & \begin{pmatrix} 0 & 0.45 \\ 0.45 & 0 \end{pmatrix} \end{array}$$



(3) (1,2) , (3,4)

(4) ((1,2),3) , 4

(5)



5. (1) μ is average of dataset. $\mu_j = \frac{1}{N}\sum_{i=1}^{N} x_{ij}$

$\|x_i-\mu\|^2 = \sum_{j=1}^{d}(x_{ij}-\mu_j)^2$

$SSE_T = \sum_{i=1}^{N}\|x_i-\mu\|^2$, and we know $SSE_T = \sum_{j=1}^{d} SSE_j$

$SSE_j = \sum_{i=1}^{N}(x_{ij}-\mu_j)^2 = \sum_{i=1}^{N} x_{ij}^2 - 2\mu_j\sum_{i=1}^{N}x_{ij} + N\mu_j^2$

$= \sum_{i=1}^{N} x_{ij}^2 - 2\mu_j\cdot N\mu_j + N\mu_j^2 = \sum_{i=1}^{N} x_{ij}^2 - N\mu_j^2$

(2) $SSE^{(j)} = \sum_{i\in C_j}\sum_{k=1}^{d}(x_{ik}-\mu_{jk})^2 = \sum_{i\in C_j}\sum_{k=1}^{d}(x_{ik}^2 - 2x_{ik}\mu_{jk} + \mu_{jk}^2)$

$= \sum_{i\in C_j}\sum_{k=1}^{d} x_{ik}^2 - n_j\sum_{k=1}^{d}\mu_{jk}^2$

(3) $SSE_T = \sum_{i=1}^{N}\sum_{k=1}^{d} x_{ik}^2 - N\sum_{k=1}^{d}\mu_k^2$

$SSE^{(1)} + SSE^{(2)} = \sum_{i=1}^{N}\sum_{k=1}^{d} x_{ik}^2 - \left(n_1\sum_{k=1}^{d}\mu_{1k}^2 + n_2\sum_{k=1}^{d}\mu_{2k}^2\right)$

$N\sum_{k=1}^{d}\mu_k^2 \le n_1\sum_{k=1}^{d}\mu_{1k}^2 + n_2\sum_{k=1}^{d}\mu_{2k}^2$

$\Rightarrow SSE_T \ge SSE^{(1)} + SSE^{(2)}$