1.1a. F
b. T
c. F
d. F
e. T

1.2

$S = \{10, 11, 2, 4, 12, 3\}$

$C_1 = \{3, 4, 11, 12\}$

$C_2 = \{2, 10\}$

a. $\mu_1 = \frac{3+4+11+12}{4} = \frac{30}{4} = \frac{15}{2} = 7.5$

b. $\mu_2 = \frac{2+10}{2} = 6$

c.

	10	11	2	4	12	3
D_1	2.5✓	3.5✓	5.5	3.5	4.5✓	4.5
D_2	4	5	4✓	2✓	6	3✓

$D_1 = \{10, 11, 12\}$, when $\mu_1 = 7.5$

$D_2 = \{2, 4, 3\}$, when $\mu_2 = 6$ (taken fr a & b)

d. $\mu_{D_1} = \frac{10+11+12}{3} = \frac{33}{3} = 11$

$\mu_{D_2} = \frac{2+4+3}{3} = 3$

e.

	10	11	12	2	4	3
E_1	1✓	0✓	1✓	9	7	8
E_2	7	8	9	1✓	1✓	0✓

E_1 and E_2 cluster membership does not change, k-means computation converges to centroids $\{11, 3\}$

2.1a. $K(x,y) = \varphi(x)^T \varphi(y) =$

$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$\begin{bmatrix} 1 & x_1^2 & \sqrt{2}x_1x_2 & x_2^2 & \sqrt{2}x_1 & \sqrt{2}x_2 \end{bmatrix} \begin{bmatrix} 1 \\ y_1^2 \\ \sqrt{2}y_1y_2 \\ y_2^2 \\ \sqrt{2}y_1 \\ \sqrt{2}y_2 \end{bmatrix}$$

Mercer's Theorem

$$= 1 + x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2 + 2x_1y_1 + 2x_2y_2$$

b. $K\left(x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, y = \begin{bmatrix} 3 \\ 4 \end{bmatrix}\right) = 1 + 1^2 \cdot 3^2 + 2 \cdot 1 \cdot 2 \cdot 3 \cdot 4 + 2^2 \cdot 4^2 + 2 \cdot 1 \cdot 3 + 2 \cdot 2 \cdot 4$

$= 1 + 10 + 48 + 64 + 6 + 16 = 59 + 86 = 145$

2.2.1

s.t. $-(d_i(w^Tx_i + b) - 1 + \xi_1) \leq 0 \mid \xi_i \geq 0$

Primal problem
$\min. \frac{1}{2}w^Tw + C\sum_{i=1}^{N} \xi_i$

$$L(w,b,\xi,\alpha) = \frac{1}{2}w^Tw + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i(d_i(w^Tx_i + b) - 1 + \xi_i) + \sum_{i=1}^{N}\beta_i(-\xi_i)$$

$$= \frac{1}{2}w^Tw + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i d_i w^T x_i - b\sum_{i=1}^{N}\alpha_i d_i + \sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\alpha_i\xi_i + \sum_{i=1}^{N}\beta_i(-\xi_i)$$

KKT conditions

$$\frac{\partial L(w,b,\xi,\alpha)}{\partial w} = \frac{1}{\cancel{2}}(\cancel{2}w) - \sum_{i=1}^{N}\alpha_i d_i x_i = 0 \qquad w = \sum_{i=1}^{N}\alpha_i d_i x_i$$

$$\frac{\partial L(w,b,\xi,\alpha)}{\partial b} = -\sum_{i=1}^{N}\alpha_i d_i = 0 \Rightarrow \sum_{i=1}^{n}\alpha_i d_i = 0$$

When $\alpha_i \geq 0$, $\beta_i \geq 0$

since $0 \leq \alpha_i \leq C$ #

$$\frac{\partial L(w,b,\xi,\alpha)}{\partial \xi} = C - \alpha_i - \beta_i = 0 \qquad \alpha_i = C - \beta_i$$

$$\frac{1}{2} w^T w = \frac{1}{2}\left[\sum_{i=1}^{N} \alpha_i d_i x_i^T\right]\left[\sum_{j=1}^{N} \alpha_j d_j x_j\right] = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j x_i^T x_j$$

$$\sum_{i=1}^{N} \alpha_i d_i w^T x_i = \sum_{i=1}^{N} \alpha_i d_i \left[\sum_{j=1}^{N} \alpha_j d_j x_j^T\right] x_i = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j x_i^T x_j$$

$$L(w, b, \alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j x_i^T x_j - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j x_i^T x_j + \sum_{i=1}^{N} \alpha_i$$

$$+ \cancelto{0}{C\sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i \xi_i - \sum_{i=1}^{N} \beta_i (\xi_i)}$$

$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j x_i^T x_j = Q(\alpha)$$

Dual problem

max. $Q(\alpha)$

s.t. $0 \le \alpha_i \le C$

$\sum_{i=1}^{N} \alpha_i d_i = 0$

2.2.2 Soft margin is used when there is no clear hyperplane $w^T x + b = 0$ that separate the 2 classes, i.e. some points are wrongly misclassified. This suggests that the data set is not linearly classifiable and the slack variable ξ in soft margin is used in minimizing classification error.