1.1a $K(x,y) = \varphi^T(x)\,\varphi(y)$

$$= \begin{pmatrix} 1 \\ x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \end{pmatrix}^T \begin{pmatrix} 1 \\ y_1^2 \\ \sqrt{2}y_1y_2 \\ y_2^2 \\ \sqrt{2}y_1 \\ \sqrt{2}y_2 \end{pmatrix}$$

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

1.1b $K(x,y) = 1 + (1)^2(3)^2 + 2(1)(2)(3)(4) + (2)^2(4)^2 + 2(1)(3) + 2(2)(4)$

$$= 144$$

1.2 1 primal problem

find : w, b

$$\min \quad f(w) = \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i$$

$$s.t. \quad d_i(w^T x_i + b) - 1 + \xi_i \geq 0,$$

$$\xi_i \geq 0$$

$$L(w, b, \xi, \alpha) = \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i \left[(d_i(w^T x_i + b) - 1 + \xi_i \right] + \sum_{i=1}^{N} \mu_i(-\xi_i)$$

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

$$- b \sum_{i=1}^{N} \alpha_i d_i + \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N} \alpha_i \xi_i + \sum_{i=1}^{N} \mu_i(-\xi_i)$$

KKT conditions

$$\frac{\partial L}{\partial w} = \frac{\partial}{\partial w} \left(\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i d_i w^T x_i \right.$$

$$\left. - b \sum_{i=1}^{N} \alpha_i d_i + \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N} \alpha_i \xi_i \right.$$

$$\left. + \sum_{i=1}^{N} \mu_i(-\xi_i) \right)$$

$$0 = \frac{1}{2} \frac{\partial (w^T w)}{\partial w} - \sum_{i=1}^{N} \alpha_i d_i \frac{\partial (w^T x_i)}{\partial w}$$

$$0 = w - \sum_{i=1}^{N} \alpha_i d_i \frac{\partial (x_i^T w)}{\partial w}$$

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

$$w = \sum_{i=1}^{N} \alpha_i d_i x_i \quad \text{— ①}$$

$$\frac{\partial L}{\partial b} = \frac{\partial}{\partial b}\left(\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i d_i w^T x_i - b \sum_{i=1}^{N} \alpha_i d_i + \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N} \alpha_i \xi_i + \sum_{i=1}^{N} \mu_i (-\xi_i) \right)$$

$$\sum_{i=1}^{N} \alpha_i d_i = 0 \quad \text{— ②}$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \mu_i$$

$$\mu_i = C - \alpha_i \quad \text{— ③}$$

$\mu_i \geq 0$: $\alpha_i \leq C$

simplify L:

① → $\frac{1}{2}w^Tw$:

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

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

① → $\sum_{i=1}^{N}\alpha_{ai} d_i w^T x_i$:

$$\sum_{i=1}^{N}\alpha_{ai} d_i w^T x_i = \sum_{i=1}^{N}\alpha_i d_i\left(\sum_{j=1}^{N}\alpha_j d_j x_j^T\right)x_i$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j d_i d_j x_i x_j^T$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j d_i d_j x_i^T x_j$$

② → $b\sum_{i=1}^{N}a_i d_i$:

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

③ → $C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\xi_i + \sum_{i=1}^{N}\mu_i(-\xi_i)$

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

$$L = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j d_i d_j x_i^T x_j + C\sum_{i=1}^{N}\xi_i + 0$$

$$-\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j d_i d_j x_i^T x_j - 0 + \sum_{i=1}^{N}\alpha_i + \sum_{i=1}^{N}\alpha_i\xi_i$$

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

$\therefore \max Q(\alpha)$

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

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

1.2 when the data set is not linearly separable, or when there are a few points in the data that is at the 'wrong' side. if hard margin is used, the resultant model will perform poorly/not general. soft margin is used to allow for some errors in prediction while making the model more general.

1.3 refer to 1.3.md

2.1

$h_\theta(x)$ means estimated probability that $y=1$ on input x.

① ✗ $P(y=0|x;\theta)$ should $= 0.65$

② ✓ correct, same as above

③ ✓ correct per definition

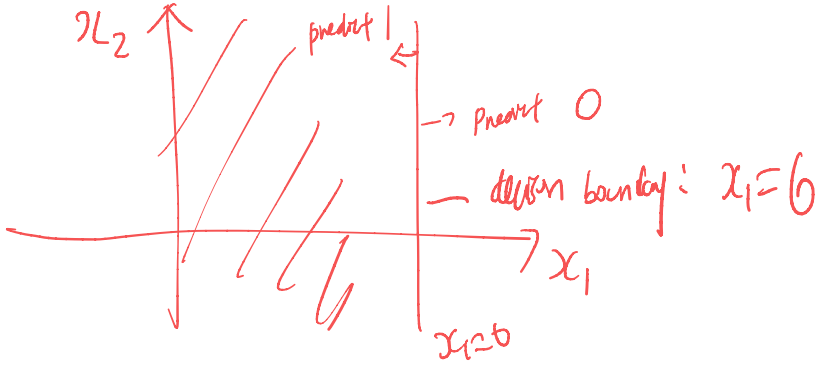
④ ✗ wrong, as above ③ is correct

2.2

$$\begin{pmatrix} 6 \\ -1 \\ 0 \end{pmatrix}^T \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} \geq 0 \Rightarrow \text{predict } 1$$

$$6 - x_1 \geq 0$$

$$x_1 \leq 6$$



2.3

decision boundary:

$$\begin{pmatrix} -9 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}^T \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = 0$$

$$-9 + x_3 + x_4 = 0$$

$$x_3 + x_4 = 9$$

2.4

1. computing log likelihood (addition) is less computationally expensive than computing likelihood (multiplication)

2. as lg of the probabilities are added together, the sum will always increase. whereas multiplying many probabilities together will result in an ever decreasing probability, approaching 0 or the limit of floating point precision.