Homework 2. Richard Pung Tuck Wei 1003523

Q1.1a)

K(x, y)

= φ(x) · φ(y)

= $[1, x_1^2, \sqrt{}2x_1x_2, x_2^2, \sqrt{}2x_1 \sqrt{}2x_2]^T *$ $[1, y_1^2, \sqrt{}2y_1y_2, y_2^2, \sqrt{}2y_1 \sqrt{}2y_2]$

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

Q1.1b)

$K([1\ 2]^T, [3\ 4]^T)$

= $1 + (1)^2 (3)^2 + 2(1)(2)(3)(4) + (2)^2(4)^2 + 2(1)(3) + 2(2)(4)$

=144

Q1.21)

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

$$\frac{\partial L(w, b, \alpha, \xi, \mu)}{\partial(w)} = 0$$

$$\frac{\partial L(w, b, \alpha, \xi, \mu)}{\partial(b)} = 0$$

$$\frac{\partial L(w, b, \alpha, \xi, \mu)}{\partial(\xi)} = 0$$

$$\frac{\partial L(w, b, \alpha, \xi, \mu)}{\partial(w)} = w - \sum_{i=1}^{N} \alpha_i d_i \left(\frac{\delta(x_i{}^T w)}{\delta w}\right)$$

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

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

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

$$\frac{\partial L(w, b, \alpha, \xi, \mu)}{\partial(\xi)} = -\alpha_i + c - \mu_i$$

$$d_i = c - \mu_i$$

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

Therefore,

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

Subject to.

$$0 \le d_i \le c$$

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

Q1.22) Soft margin is preferred when the data is not linearly separable due to some data being noisy or prone to error.

Q2.1) Our estimate for $P(y = 1|x; \theta)$ is 0.35 as h(x) predicts the probability that y =1. Our estimate for $P(y = 0|x; \theta)$ is 0.65 as $P(y = 1|x; \theta)$ is 0.35 and $P(y = 0|x; \theta) = 1 - P(y = 1|x; \theta)$ is 0.65.

Q2.2) decision boundary:

if $6 - x_1 >= 0, y = 1$

if $6 - x_1 < 0, y = 0$

Q2.3) decision boundary:

if $-9 + x_1^2 + x_2^2 >= 0, y = 1$

if $-9 + x_1^2 + x_2^2 < 0, y = 0$

Q2.4)

The derivative of the log likelihood is easier to calculate than the derivative of the likelihood function. This is because the log of the product is equivalent to the sum of log for each term.

Also, product of number may quicky converge to infinity or zero.