50.007 Machine Learning HW2

Victoria Yong 1004455

Q1.1 a) F

b) T

c) F

d) F

e) T

Q1.2 a) $\mu_1 = \frac{3+4+11+12}{4} = 7.5$

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

c) Since the two centroids are 6 and 7.5, all points less than 6 will be clustered to D_2 and all points more than 7.5 will be clustered to D_1.

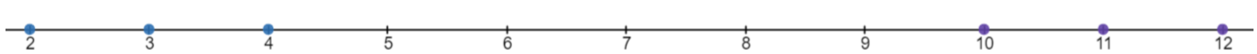
$D_1 = \{10, 11, 12\}$

$D_2 = \{2, 3, 4\}$

d) Centroid of $D_1 = \frac{10+11+12}{3} = 11$

Centroid of $D_2 = \frac{2+3+4}{3} = 3$

e) The clustering is stable. The optimal solution is one where each cluster has minimal variance. The variance of both D clusters is 0.667, which is much lower compared to the variance $C_1 = 16.25$ and $C_2 = 16$. This is the lowest possible variance for the clusters given that the data are all different integer values. Calculating the new centroids of this cluster would also yield the same results. Furthermore, considering the values of the data points, it is also easy to visually cluster them by plotting the points on a number line as shown below.



Q2.1 a) $K(x_1 y) = \varphi(x)^T \varphi(y)$

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

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

b) $K(x_1 y) = \varphi([1\ \ 2])^T \varphi([3\ \ 4])$

$$= 1 + (1^2)(3^2) + 2(1)(2)(3)(4) + 2^2(4^2) + 2(1)(3) + 2(2)(4)$$
$$= 144$$

Q2.2 a)

1. $Primal\ Problem: minimize\ \frac{1}{2}\omega^T\omega + c\sum_{i=1}\varepsilon_i$

   $-d_i(\omega^T x_i + b) + 1 - \varepsilon_i \leq 0$

2. $L(\omega, b, \varepsilon, a) = \frac{1}{2}\omega^T\omega + c\sum_{i=1}^{N}\varepsilon_i - \sum_{i=1}^{N}\alpha_i\,[d_i(\omega^T x_i + b) - 1 + \varepsilon_i] + \sum_{i=1}^{N}\mu_i(-\Sigma_i)$

   $$= \frac{1}{2}\omega^T\omega + c\sum_{i=1}^{N}\varepsilon_i - \sum_{i=1}^{N}\alpha_i d_i\omega^T x_i - b\sum_{i=1}^{N}\alpha_i d_i + \sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\alpha_i\varepsilon_i + \sum_{i=1}^{N}\mu_i(-\varepsilon_0)$$

3. $\frac{\partial L}{\partial w} = \omega - \sum_{i=1}^{N}\alpha_i d_i x_i = 0$

   $\omega = \sum_{i=1}^{N}\alpha_i d_1 x_i$

   $\frac{\partial L}{\partial b} = -\sum_{i=1}^{N}\alpha_i d_i = 0$

   $\frac{\partial L}{\partial \varepsilon_i} = c - \alpha_i - \mu_i = 0$

   $c = \alpha_i + \mu_i$

   $KKT\ Conditions:$

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

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

   $c - \alpha_i d_i = 0$

   $d_i(\omega^T x_i + b) - 1 + \varepsilon_i \geq 0$

   $\alpha_i(d_1(\omega^T x_i + b) - 1 + \varepsilon_i) = 0$

   $\alpha_i \geq 0, \mu_i \geq 0, \mu_i\varepsilon_i = 0$

4. $Simplify\ L,\ \frac{1}{2}\omega^T\omega = \frac{1}{2}\sum_{i=1}^{N}\alpha_i d_i x_i^T\sum_{k=1}^{N}\alpha_k d_k x_k$

   $$= \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{N}\alpha_i\alpha_k d_i d_k x_i^T x_k$$

   $$\sum_{i=1}^{N}\alpha_i d_i\omega^T x_i = \sum_{i=1}^{N}\sum_{k=1}^{N}\alpha_i\alpha_k d_i d_k x_i^T x_k$$

   $$L(\omega, b, \varepsilon, \alpha) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{N}\alpha_i\alpha_k d_i d_k x_i^T x_k + \sum_{i=1}^{N}\alpha_i$$

5. $Dual\ Problem:$

   $maxmise\ L(\omega, b, \varepsilon, \alpha) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{N}\alpha_i\alpha_k d_i d_k x_i^T x_k + \sum_{i=1}^{N}\alpha_i$

   $s.t.\ \sum_{i=1}^{N}\alpha_i d_i = 0\ , 0 \leq \alpha_i \leq c$

b) A soft margin is preferable when data is not linearly separable. This allows a degree of misclassification which would prevent overfitting of the data.

Q2.3 See attached script svmtest.py for code.

The best kernel is the Radial Basis Function kernel with an accuracy of 87.3%.

```
.....*.*
optimization finished, #iter = 1012
nu = 0.395283
obj = -47.881165, rho = 1.910755
nSV = 76, nBSV = 44
Total nSV = 76
Accuracy = 79.3651% (50/63) (classification)
Linear Kernel Accuracy: (79.36507936507937, 0.8253968253968254, 0.3434353678775586)
*
optimization finished, #iter = 80
nu = 0.937931
obj = -126.052163, rho = 0.969786
nSV = 139, nBSV = 132
Total nSV = 139
Accuracy = 55.5556% (35/63) (classification)
Polynomial Kernel Accuracy: (55.55555555555556, 1.7777777777777777, 0.018909899888765295)
*
optimization finished, #iter = 74
nu = 0.789215
obj = -94.116556, rho = 0.244822
nSV = 117, nBSV = 108
Total nSV = 117
Accuracy = 87.3016% (55/63) (classification)
RBF Kernel Accuracy: (87.3015873015873, 0.5079365079365079, 0.5545131845841785)
*
optimization finished, #iter = 83
nu = 0.839768
obj = -106.649422, rho = 1.168119
nSV = 126, nBSV = 118
Total nSV = 126
Accuracy = 82.5397% (52/63) (classification)
Sigmoid Kernel Accuracy: (82.53968253968253, 0.6984126984126984, 0.43500807311226347)

[Done] exited with code=0 in 1.699 seconds
```