Homework 5

B99202064 Kuan Hou Chan

1 Ans: (C)

a quadratic programming problem with N size of data set and d dimension input space will have d+1 variables and N constraints.

2 Ans: C

By using cvxopt package, we can find the optimization of $b = -9, w_1 = 2, w_2 = 0$
Hence the equation will be $2z_1 - 9 = 0 \Rightarrow$ $z_1 = 4.5$

3 Ans: B,D

By using cvxopt package, we obtain that
$\alpha_1 = 6.69 * 10^{-9}, \alpha_2 = 0.7, \alpha_3 = 0.7, \alpha_4 = 0.89, \alpha_5 = 0.26, \alpha_6 = 0.26, \alpha_7 = 6.4 * 10^{-10}$
Hence, $\sum_{n=1}^{7} \alpha_n \approx 2.8148$ and $min_{1 \le n \le 7} \alpha_n = \alpha_7$

4 Ans: B

By using Q3 answer, first we can obtain $b^* = \sum_{\alpha_n > 0} y_n * \alpha_n * (1 + x_n * x_{sv})^2 = -1.67$
Then $g = sign(\sum_{\alpha_n > 0} y_n * \alpha_n^* K(x_n, X) + b)$
$= -0.7 * [1 + (0,1)^T(x1, x2)]^2 - 0.7 * [1 + (0,-1)^T(x1, x2)]^2 + 0.89 * [1 + (-1,0)^T(x1, x2)]^2 + 0.26 * [1 + (0,2)^T(x1, x2)]^2 + 0.26 * [1 + (0,-2)^T(x1, x2)]^2$
$= -1.67 + 0.6x_2^2 - 1.78x_1 + 0.89x_1^2$
Hence the choice B is the same with the answer we obtain.

5 Ans: (C)

Although Q2 and Q4 both are polynomial transform, but the transform is different in Z space,so the curve should be different in X space.

6 Ans: D

We can easily write down the dual problem that is $L(R, c, \lambda) = R^2 + \sum_{n=1}^{N} \lambda_n (||x_n - c||^2 - R^2)$

7 Ans: A,C,D

By KKT conditions 3, $\nabla_{(R,c)} L(R, c, \lambda) = 0$
$\Rightarrow \frac{\partial L}{\partial R} = 2R(1 - \sum_{n=1}^{N} \lambda_n) = 0$, if $R \neq 0, \sum_{n=1}^{N} \lambda_n = 1$,A is true.
$\Rightarrow \frac{\partial L}{\partial c} = \sum_{n=1}^{N} \lambda_n [2(x_n - c)(-1)] = 0$,if $\sum_{n=1}^{N} \lambda_n \neq 0$,$c = \frac{\sum_{n=1}^{N} \lambda_n x_n}{\sum_{n=1}^{N} \lambda_n}$,D is true
By KKT conditions 2, $\sum_{n=1}^{N} \lambda_n (||x_n - c||^2 - R^2) = 0$,if $\sum_{n=1}^{N} \lambda_n = 0$ then $||x_n - c||^2 - R^2 \neq 0$, B is false.
And also if $||x_n - c||^2 - R^2 < 0, \Rightarrow \sum_{n=1}^{N} \lambda_n = 0$,C is true.

8 Ans: A

$L(R,c,\lambda) = R^2 + \sum_{n=1}^{N}\lambda_n(||x_n - c||^2 - R^2) = R * R(1 - \sum_{n=1}^{N}\lambda_n) + \sum_{n=1}^{N}\lambda_n((||x_n - c||^2))$
By using the answer of last question, $\sum_{n=1}^{N}\lambda_n = 1, R(1 - \sum_{n=1}^{N}\lambda_n) = 0$
$\Rightarrow L = 0 + \sum_{n=1}^{N}\lambda_n(||x_n - c||^2) = \sum_{n=1}^{N}\lambda_n(||x_n - \frac{\sum_{m=1}^{M}\lambda_m x_m}{\sum_{m=1}^{M}\lambda_m}||^2)$
$= \sum_{n=1}^{N}\lambda_n(||x_n - \sum_{m=1}^{M}\lambda_m x_m||^2)$

9 Ans: C

Expand $L = \sum_{n=1}^{N}\lambda_n(||z_n - \sum_{m=1}^{M}\lambda_m z_m||^2)$
obtain $L = \sum_{n=1}^{N}\lambda_n([z_n - \sum_{m=1}^{M}\lambda_m z_m]^T[z_n - \sum_{m=1}^{M}\lambda_m z_m])$
$L = \sum_{n=1}^{N}\lambda_n[z_n^T z_n - 2\sum_{m=1}^{M}\lambda_m z_n^T z_m + \sum_{m=1}^{M}\lambda_m\sum_{m'=1}^{M'}\lambda'_m z_m^T z_{m'}]$
$L = \sum_{n=1}^{N}\lambda_n K(x_n,x_n) - 2\sum_{n=1}^{N}\lambda_n\sum_{m=1}^{M}\lambda_m K(x_n,x_m) + \sum_{n=1}^{N}\lambda_n\sum_{m=1}^{M}\lambda_m\sum_{m'=1}^{M'}\lambda'_m K(x_m,x_{m'})]$
Since $\sum_{n=1}^{N}\lambda_n = 1, \therefore$
$L = \sum_{n=1}^{N}\lambda_n K(x_n,x_n) - 2\sum_{n=1}^{N}\lambda_n\sum_{m=1}^{M}\lambda_m K(x_n,x_m) + 1 * \sum_{m=1}^{M}\lambda_m\sum_{m'=1}^{M'}\lambda'_m K(x_m,x_{m'})]$
$L = \sum_{n=1}^{N}\lambda_n K(x_n,x_n) - 1\sum_{n=1}^{N}\lambda_n\sum_{m=1}^{M}\lambda_m K(x_n,x_m)$

10 Ans: A

By using KKT conditions 2 that if $\lambda_n \neq 0, ||z_n - c||^2 - R^2 = 0$
then $R^2 = ||z_n - \sum_{m=1}^{M}\lambda_m z_m||^2 = z_i^T z_i - 2z_i^T\sum_{m=1}^{M}\lambda_m z_m + \sum_{n=1}^{N}\lambda_n\sum_{m=1}^{M}\lambda_m z_n^T z_m$
$R = \sqrt{K(x_i,x_i) - 2\sum_{m=1}^{M}\lambda_m K(x_i,x_m) + \sum_{n=1}^{N}\lambda_n\sum_{m=1}^{M}\lambda_m K(x_n,x_m)}$

11 Ans: D

Q11QQ
Q11QQ
Q11QQ
Q11QQ
Q11QQ

12 Ans: A,C,D

Since K_1 and K_2 are valid kernels means that they are positive definite, $x^T K_1 x \geq 0, x^T K_2 x \geq 0$
A: $K(x,x') = \phi(x)\phi(x') = (\phi_1(x),\phi_2(x))(\phi_1(x'),\phi_2(x')) = \phi_1(x)^T\phi_1(x') + \phi_2(x)^T\phi_2(x') = K_1(x,x') + K_2(x,x')$
B: if $K_1(x,x') < K_2(x,x') \Rightarrow K(x,x') = K_1(x,x') - K_2(x,x') < 0$ then $K(x,x')$ not PD matrix and will become invalid

C: $K(x,x') = \phi(x)\phi(x') = \sum_m \phi_m(x)\phi_m(x') = \sum_i\sum_j \phi_{1i}(x)^T\phi_{2j}(x)\phi_{1i}(x')^T\phi_{2j}(x')$
$= \sum_i(\phi_{1i}(x)^T\phi_{1i}(x'))\sum_j(\phi_{2j}(x)^T\phi_{2j}(x')) = \phi_1(x)^T\phi_1(x')\phi_2(x)^T\phi_2(x') = K_1(x,x')K_2(x,x')$
D: Since $K_1/K_2 = K_1K_2^{-1}$, $\because$ inverse of PD matrix still is PD matrix,and 2 PD matrix product still is PD matrix, hence K is valid kernel.

13 Ans: B,D

A: Use the result from last question of B, then K will also become invalid kernel.
B: Valid. $K(x,x') = \phi(x)\phi(x') = (\sqrt{f}\phi_1(x))(\sqrt{f}\phi_1(x')) = f\phi_1(x)\phi_1(x') = fK_1(x,x')$
C: Use Taylor expansion of $e^{-x} = \lim_{x\to 0}(1 - x + \frac{x^2}{2!} - .... + ...)$, by Q12 of B,if we take x as K, then the result will be

invalid kernel.
D: if $0 < K_1(x, x^{'}) < 1$, then the diagonal term of matrix will not occur 0 then $K(x, x^{'})$ still is a PD and valid kernel.

14 Ans:

Since

$$b = y_s - \sum_n \alpha_n y_n (K(x_n, x_s))$$

$$g_{svm} = sign(\sum_n \alpha_n y_n K(x_n, x) + b)$$

$$\tilde{b} = y_s - \sum_n \tilde{\alpha_n} y_n (pK(x_n, x_s) + q)$$

$$\tilde{g}_{svm} = sign(\sum_n \tilde{\alpha_n} y_n (pK(x_n, x_s) + q)$$

$$\Rightarrow \tilde{g}_{svm} = \sum_n y_n p \tilde{\alpha_n} K(x_n, x) - \sum_n y_n p \tilde{\alpha_n} K(x_n, x_s)$$

compared $\tilde{g}_{svm} = g_{svm} \Rightarrow \tilde{\alpha} = \frac{\alpha}{p}$
By kKT conditions, $(C - \alpha_n^*)\xi_n^* = 0$ and if $\xi_n^* \neq 0, C = \alpha_n^*$
Then $\tilde{C} = \frac{C}{p}$

15 Bonus 21

If no free SV exist, it cannot conclude that the data is not linearly separable because in this case maybe $\alpha = C or \alpha > C$ implies that maybe we already set a large error tolerance or C is too small to cover all the data become bounded SV or non SV
Hence we cannot conclude the data whether or not is linearly separable.

16 Bonus 22

Again, like the last question, but plus the condition of $\xi_n > 1$, then it is means that the bounded SV can have large violation on the boundary,
Hence we can conclude that the data is not linearly separable.