EE E6887 (Statistical Pattern Recognition)
Solutions for homework 6

P.1 In the lecture notes, we have formulated the unconstrained Lagrangean as follows

$$L_p = \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{l} \alpha_i \Big(y_i(\mathbf{w}^t \mathbf{x}_i + b) - 1 \Big)$$

subject to $\alpha_i \geq 0$. This is called the primal form.

Take the derivatives of the above with respect to $\mathbf{w}$ and b. By making the derivatives vanish, show that you can derive the following "dual form":

$$L_D = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j$$

Answer:

$$\begin{aligned} L_p &= \frac{1}{2}\mathbf{w}^t\mathbf{w} - \sum_{i=1}^{l} \alpha_i \Big(y_i(\mathbf{w}^t \mathbf{x}_i + b) - 1 \Big) \\ &= \frac{1}{2}\mathbf{w}^t\mathbf{w} - \sum_{i=1}^{l} \alpha_i y_i(\mathbf{w}^t \mathbf{x}_i + b) + \sum_{i=1}^{l} \alpha_i \end{aligned}$$

$$\begin{aligned} \frac{\partial L_p}{\partial \mathbf{w}} &= \mathbf{w} - \sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i \\ \frac{\partial L_p}{\partial b} &= \sum_{i=1}^{l} \alpha_i y_i \end{aligned}$$

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0 \;\;\Rightarrow\;\; \mathbf{w} = \sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i$$

$$\frac{\partial L_p}{\partial b} = 0 \;\;\Rightarrow\;\; \sum_{i=1}^{l} \alpha_i y_i = 0$$

Substitute them into the original primal form, we get

$$\begin{aligned}
L_p &= \frac{1}{2}\sum_{i=1}^{l}\alpha_i y_i(\mathbf{x}_i)^t\sum_{j=1}^{l}\alpha_j y_j\mathbf{x}_j - \sum_{i=1}^{l}\alpha_i y_i(\sum_{j=1}^{l}\alpha_j y_j(\mathbf{x}_j)^t\mathbf{x}_i + b) + \sum_{i=1}^{l}\alpha_i \\
&= \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i y_i\alpha_j y_j\mathbf{x}_i\cdot\mathbf{x}_j - \sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i y_i\alpha_j y_j\mathbf{x}_i\cdot\mathbf{x}_j - b\sum_{i=1}^{l}\alpha_i y_i + \sum_{i=1}^{l}\alpha_i \\
&= \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j\mathbf{x}_i\cdot\mathbf{x}_j = L_D
\end{aligned}$$

P.2 Repeat Example 2 using the same $\psi(\cdot)$ but with the following four points:

$$\begin{aligned}
\omega_1 &: (1,5)^t \quad (-2,-4)^t \\
\omega_2 &: (2,3)^t \quad (-1,5)^t
\end{aligned}$$

Note you need to find the Lagrange multipliers α_i, point out which samples are support vectors, derive the discriminant function, and derive the equation of classification hyperplane in the higher-dimensional space. Though it is not mandatory, you are encouraged to plot the decision hyperplane and the hyperplanes crossing the support vectors in the original space ($\mathbf{x}_1$, $\mathbf{x}_2$).

Answer: Assume $\mathbf{x}_1 = (1,5)^t$, $\mathbf{x}_2 = (-2,-4)^t$, $\mathbf{x}_3 = (2,3)^t$, $\mathbf{x}_4 = (-1,5)^t$. $y_1 = 1$, $y_2 = 1$, $y_3 = -1$, $y_4 = -1$. The dual form in the transformed space is given by:

$$\begin{aligned}
L_D &= \sum_{i=1}^{4}\alpha_i - \frac{1}{2}\sum_{i=1}^{4}\sum_{j=1}^{4}\alpha_i\alpha_j y_i y_j\varphi(\mathbf{x}_i)\cdot\varphi(\mathbf{x}_j) \\
&= \sum_{i=1}^{4}\alpha_i - \frac{1}{2}\alpha_1^2\varphi^2(\mathbf{x}_1) - \frac{1}{2}\alpha_2^2\varphi^2(\mathbf{x}_2) - \frac{1}{2}\alpha_3^2\varphi^2(\mathbf{x}_3) - \frac{1}{2}\alpha_4^2\varphi^2(\mathbf{x}_4) \\
&\quad -\alpha_1\alpha_2\varphi(\mathbf{x}_1)\cdot\varphi(\mathbf{x}_2) + \alpha_1\alpha_3\varphi(\mathbf{x}_1)\cdot\varphi(\mathbf{x}_3) \\
&\quad +\alpha_1\alpha_4\varphi(\mathbf{x}_1)\cdot\varphi(\mathbf{x}_4) + \alpha_2\alpha_3\varphi(\mathbf{x}_2)\cdot\varphi(\mathbf{x}_3) \\
&\quad +\alpha_2\alpha_4\varphi(\mathbf{x}_2)\cdot\varphi(\mathbf{x}_4) - \alpha_3\alpha_4\varphi(\mathbf{x}_3)\cdot\varphi(\mathbf{x}_4)
\end{aligned}$$

under the constraint

$$\sum_{i=1}^{4} \alpha_i y_i = 0$$
$$\alpha_i \geq 0, \quad i = 1, \ldots, 4$$

So we have

$$\alpha_1 + \alpha_2 = \alpha_3 + \alpha_4$$
$$\alpha_i \geq 0, \quad i = 1, \ldots, 4$$

And we choose $\varphi(\mathbf{x}_i) = [1, \sqrt{2}x_{i1}, \sqrt{2}x_{i2}, \sqrt{2}x_{i1}x_{i2}, (x_{i1})^2, (x_{i2})^2]^t$, so

$$\begin{aligned}
\varphi(\mathbf{x}_1) &= [1, \sqrt{2}, 5\sqrt{2}, 5\sqrt{2}, 1, 25]^t \\
\varphi(\mathbf{x}_2) &= [1, -2\sqrt{2}, -4\sqrt{2}, 8\sqrt{2}, 4, 16]^t \\
\varphi(\mathbf{x}_3) &= [1, 2\sqrt{2}, 3\sqrt{2}, 6\sqrt{2}, 4, 9]^t \\
\varphi(\mathbf{x}_4) &= [1, -\sqrt{2}, 5\sqrt{2}, -5\sqrt{2}, 1, 25]^t
\end{aligned}$$

$$\begin{aligned}
L_D &= \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{729}{2}\alpha_1^2 - \frac{441}{2}\alpha_2^2 - \frac{196}{2}\alpha_3^2 - \frac{729}{2}\alpha_4^2 \\
&\quad -441\alpha_1\alpha_2 + 324\alpha_1\alpha_3 + 625\alpha_1\alpha_4 + 225\alpha_2\alpha_3 + 289\alpha_2\alpha_4 - 196\alpha_3\alpha_4
\end{aligned}$$

and our problem turns to maximize L_D

$$L_D = [1, 1, 1, 1] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} - \frac{1}{2} [\alpha_1, \alpha_2, \alpha_3, \alpha_4] \begin{bmatrix} 729 & 441 & -324 & -625 \\ 441 & 441 & -225 & -289 \\ -324 & -225 & 196 & 196 \\ -625 & -289 & 196 & 729 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}$$

under constrains

$$\alpha_1 + \alpha_2 - \alpha_3 - \alpha_4 = 0$$
$$\alpha_i \geq 0, \quad i = 1, \ldots, 4$$

let

$$Y = \begin{bmatrix} 729 & 441 & -324 & -625 \\ 441 & 441 & -225 & -289 \\ -324 & -225 & 196 & 196 \\ -625 & -289 & 196 & 729 \end{bmatrix}, \quad u = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}, \quad a = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

L_D can be written as:

$$L_D = a^t u - \frac{1}{2} u^t Y u$$

Use the lagrange multiplier technique for the constrained optimization problem (refer to `http://www.cs.ubc.ca/spider/ascher/542/chap9.pdf`), we get

$$J_D = a^t u - \frac{1}{2} u^t Y u + \lambda c^t u$$

where $\lambda \geq 0$, $c = [1, 1, -1, -1]^t$.

So we take derivative as follows:

$$\begin{aligned} \nabla_u J_D &= a - Yu + \lambda c = 0 \\ \frac{\partial J_D}{\partial \lambda} &= c^t u = 0 \\ \Rightarrow \quad \begin{bmatrix} Y & -c \\ c^t & 0 \end{bmatrix} \begin{bmatrix} u \\ \lambda \end{bmatrix} &= \begin{bmatrix} a \\ 0 \end{bmatrix} \end{aligned}$$

So

$$\begin{bmatrix} u \\ \lambda \end{bmatrix} = \begin{bmatrix} Y & -c \\ c^t & 0 \end{bmatrix}^{-1} \begin{bmatrix} a \\ 0 \end{bmatrix} = \begin{bmatrix} 0.0154 \\ 0.0067 \\ 0.0126 \\ 0.0095 \\ 3.1713 \end{bmatrix}$$

That is:

$$\begin{aligned} \alpha_1 &= 0.0154 \\ \alpha_2 &= 0.0067 \\ \alpha_3 &= 0.0126 \\ \alpha_4 &= 0.0095 \end{aligned}$$

Since all $\alpha_i \neq 0$, all of the samples are support vectors, and the discriminant function in the transformed space is $g(\varphi(\mathbf{x})) = \mathbf{w}^t \varphi(\mathbf{x}) + \tilde{b}$, where

$$\mathbf{w} = \sum_{i=1}^{4} \alpha_i y_i \varphi(\mathbf{x}_i)$$

In the original space, the discriminant function is then given by

$$g(\mathbf{x}) = \sum_{i=1}^{4} \alpha_i \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}) + \tilde{b}$$

where

$$\begin{aligned} \tilde{b} &= \frac{1}{4} \sum_{i=1}^{4} \left[y_i - \sum_{j=1}^{4} \alpha_j y_j \varphi(\mathbf{x}_j) \cdot \varphi(\mathbf{x}_i) \right] \\ &= -3.1647 \end{aligned}$$

So finally the discriminant function is In the original space, the discriminant function is then given by

$$g(\mathbf{x}) = -0.0274x_1 - 0.0702x_2 + 0.1648x_1x_2 - 0.0177x_1^2 - 0.2283x_2^2 - 3.1647$$