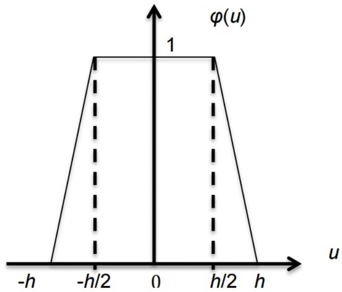
HW6

1. We perform Parzen Window density estimation, using trapezoidal window functions given (in unnormalized form) in the figure below: (15 pts)



Choose $h = 1$. Assume that we have the following data

$$\mathcal{D}_{\omega_1} = \{0, 2, 5\}$$
$$\mathcal{D}_{\omega_2} = \{4, 7\}$$

(a) Sketch or plot the Parzen window estimates of the pdfs $p(x|\omega_1)$ and $p(x|\omega_2)$. Please label pertinent values on both axes.

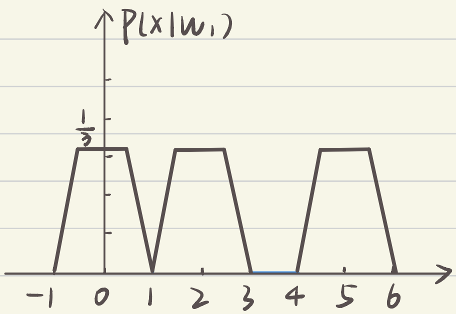
(a)

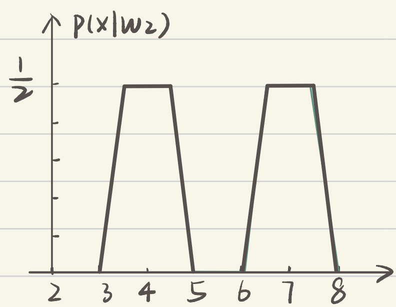
$$\phi(u) = \begin{cases} 2|u|+1, & -1 \le u < -0.5 \\ 1, & -0.5 \le u < 0.5 \\ -2|u|+1, & 0.5 \le u \le 1 \\ 0, & \text{otherwise} \end{cases}$$

$\because p(x|w) = \frac{1}{n} \sum_{i=1}^{n} \phi(\frac{x - x_i}{n})$

$\therefore$ for $D_{w_1} = \{0, 2, 5\} \Rightarrow p(x|w_1) = \frac{1}{3}[\phi(x-0) + \phi(x-2) + \phi(x-5)]$

$D_{w_2} = \{4, 7\} \Rightarrow p(x|w_2) = \frac{1}{2}[\phi(x-4) + \phi(x-7)]$





3. Consider the following training data set:

$$\mathbf{x}_1 = [1, 0]^T, z_1 = -1$$
$$\mathbf{x}_2 = [0, 1]^T, z_2 = -1$$
$$\mathbf{x}_3 = [0, -1], z_3 = -1$$
$$\mathbf{x}_4 = [-1, 0]^T, z_4 = 1$$
$$\mathbf{x}_5 = [0, 2]^T, z_5 = 1$$
$$\mathbf{x}_6 = [0, -2]^T, z_6 = 1$$
$$\mathbf{x}_7 = [-2, 0]^T, z_7 = 1$$

Use following nonlinear transformation of the input vector $\mathbf{x} = [x_1, x_2]^T$ to the transformed vector $\mathbf{u} = [\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x})]^T$: $\varphi_1(\mathbf{x}) = x_2^2 - 2x_1 + 3$ and $\varphi_2(\mathbf{x}) = x_1^2 - 2x_2 - 3$.

What is the equation of the optimal separating "hyperplane" in the $\mathbf{u}$ space? (15 pts)

$U_1 = [\varphi_1(X_1), \varphi_2(X_1)]^T = [0^2 - 2 + 3, 1^2 - 0 - 3]^T = [1, -2]^T$

$U_2 = [\varphi_1(X_2), \varphi_2(X_2)]^T = [1^2 - 0 + 3, 0^2 - 2 - 3]^T = [4, -5]^T$

Similarly: $U_3 = [4, -1]^T$, $U_4 = [5, -2]^T$, $U_5 = [7, -7]^T$

$U_6 = [7, 1]^T$, $U_7 = [7, 1]^T$

$\therefore$ for transformed vector $U = [u_1, u_2]^T$

$$\begin{cases} z = -1: [1, -2]^T, [4, -5]^T, [4, -1]^T & \Rightarrow u_1 \le 4 \\ z = 1: [5 -2]^T, [7, -7]^T, [7, 1]^T, [7, 1]^T & \Rightarrow u_1 \ge 5 \end{cases}$$

$\therefore$ $u_1 = 4.5$ can seperate two group

$\therefore$ we can find a seperating hyperplane

$1 \cdot u_1 + 0 \cdot u_2 = 4.5$

$\Rightarrow [1, 0] \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 4.5$

$\Rightarrow w^T u + b = 0$ where $w = [0, 1]^T$, $b = -4.5$

4. Consider the following training data set : (25 pts)

$$\mathbf{x}_1 = [0, 0]^T, z_1 = -1$$
$$\mathbf{x}_2 = [1, 0]^T, z_2 = 1$$
$$\mathbf{x}_3 = [0, -1], z_3 = 1$$
$$\mathbf{x}_4 = [-1, 0]^T, z_4 = 1$$

Note that in the following, you need to use equations that describe $\mathbf{w}$ and give rise to the dual optimization problem.

(a) Write down the dual optimization problem for training a Support Vector Machine with this data set using the polynomial kernel function

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + 1)^2$$

(b) Solve the optimization problem and find the optimal λ_i's using results about quadratic forms and check the results with Wolfram Alpha or any software package.

(a) $\because k(X_i, X_j) = (X_i^T X_j + 1)^2$

$\therefore$ dual problem

$$\max_{\lambda} \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j z_i z_j (X_i^T X_j + 1)^2$$

s.t. $\sum_{i=1}^{N} \lambda_i z_i = -\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 0 \Rightarrow \lambda_1 = \lambda_2 + \lambda_3 + \lambda_4$ $\quad \lambda_i \geq 0$

(b) Let $\lambda = [\lambda_1, \lambda_2, \lambda_3, \lambda_4]^T$

$\Rightarrow \max_{\lambda} 1^T \lambda - \frac{1}{2} \lambda^T Q \lambda$ where $Q_{ij} = z_i z_j k(X_i, X_j)$

$\because k(X_1, X_1) = (X_1^T X_1 + 1)^2 = (0+1)^2 = 1$

$k(X_1, X_2) = (X_1^T X_2 + 1)^2 = (0+1)^2 = 1 = k(X_2, X_1)$

$\therefore$ Similarly $k(X_2, X_2) = 4$ $\quad K(X_3, X_3) = 4$ $\quad k(X_4, X_4) = 4$

$k(X_1, X_3) = k(X_3, X_1) = 1$, $\quad k(X_1, X_4) = k(X_4, X_1) = 1$

$k(X_2, X_3) = k(X_3, X_2) = 1$, $\quad k(X_2, X_4) = k(X_4, X_2) = 0$

$k(X_3, X_4) = k(X_4, X_3) = 1$

$\therefore$ kernal matrix

$$K = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 4 & 1 & 0 \\ 1 & 1 & 4 & 1 \\ 1 & 0 & 1 & 4 \end{bmatrix}$$

$\therefore$ let $Z = [-1, 1, 1, 1]^T$

$$\therefore Q = (ZZ^T) \circ K = \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & 1 & 1 \end{bmatrix} \circ \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 4 & 1 & 0 \\ 1 & 1 & 4 & 1 \\ 1 & 0 & 1 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \end{bmatrix} \circ \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 4 & 1 & 0 \\ 1 & 1 & 4 & 1 \\ 1 & 0 & 1 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 4 & 1 & 0 \\ -1 & 1 & 4 & 1 \\ -1 & 0 & 1 & 4 \end{bmatrix}$$

$$\therefore \max_{\lambda} \ \lambda_1+\lambda_2+\lambda_3+\lambda_4 - \frac{1}{2}[\lambda_1\ \lambda_2\ \lambda_3\ \lambda_4]\, Q\, [\lambda_1\ \lambda_2\ \lambda_3\ \lambda_4]^T$$

$$\text{s.t. } \lambda_1 = \lambda_2+\lambda_3+\lambda_4, \quad \lambda_i \geq 0$$

$$= \max \ \lambda_1+\lambda_2+\lambda_3+\lambda_4 - \frac{1}{2}(\lambda_1^2+4\lambda_2^2+4\lambda_3^2+4\lambda_4^2+2(-\lambda_1\lambda_2-\lambda_1\lambda_3-\lambda_1\lambda_4+\lambda_2\lambda_3+\lambda_3\lambda_4))$$

$\because X_2 = [1, 0]^T$ and $X_4 = [-1, 0]^T$ are symmetric

$\therefore$ Suppose $\lambda_2 = \lambda_4 = a$, $\lambda_3 = b \Rightarrow \lambda_1 = 2a+b$

$$\therefore \max_{\lambda} \ (2a+b)+a+b+a - \frac{1}{2}(4a^2+4ab+b^2+4a^2+4b^2+4a^2+ 2(-2a^2-ab-2ab-b^2-2a^2-ab+ab+ab))$$

$$= \max_{\lambda} \ 4a+2b-2a^2-\frac{3}{2}b^2$$

$\therefore$ Let $f(a,b) = -2a^2 - \frac{3}{2}b^2 + 4a + 2b$

$$\therefore \text{ Let } \begin{cases} \frac{\partial f}{\partial a} = -4a + 4 = 0 \\ \frac{\partial f}{\partial b} = -3b + 2 = 0 \end{cases} \Rightarrow \begin{cases} a = 1 \\ b = \frac{2}{3} \end{cases}$$

$\therefore \lambda_1 = 2a + b = \frac{8}{3}, \quad \lambda_2 = 1, \quad \lambda_3 = \frac{2}{3}, \quad \lambda_4 = 1$

using python we can check the result is correct

q4.py > ...

```python
import numpy as np
from scipy.optimize import minimize

K = np.array([
    [1, 1, 1, 1],
    [1, 4, 1, 0],
    [1, 1, 4, 1],
    [1, 0, 1, 4]
], dtype='double')

z = np.array([-1, 1, 1, 1])
Q = np.outer(z, z) * K

def objective(l):
    return -np.sum(l) + 0.5 * np.dot(l, np.dot(Q, l))

constraints = {'type': 'eq', 'fun': lambda l: np.dot(l, z)}
bounds = [(0, None)] * 4
l0 = np.zeros(4)

res = minimize(objective, l0, bounds=bounds, constraints=constraints)
print(res.x)
```

问题　输出　调试控制台　终端　端口

```
[2.66668028 0.99997915 0.66672111 0.99998002]
```

$\approx \frac{8}{3} \qquad \approx 1 \qquad \approx \frac{2}{3} \qquad \approx 1$

(c) Show that the equation of the decision boundary in a kernel SVM $\mathbf{w}^T\mathbf{u} + w_0 = 0$ can be represented as $g(\mathbf{x}) = \sum_{i=1}^{N} \lambda_i z_i \kappa(\mathbf{x}_i, \mathbf{x}) + w_0$.

(c) $\because$ in SVM, decision boundary is $w^T u + w_0 = 0$

where u corresponds to the vector after mapping $\phi(x)$

$\because w = \sum_{i=1}^{N} \lambda_i z_i u_i$

$\therefore w^T u + w_0 = (\sum_{i=1}^{N} \lambda_i z_i u_i)^T u + w_0$

$= \sum_{i=1}^{N} \lambda_i z_i u_i^T u + w_0$

$= \sum_{i=1}^{N} \lambda_i z_i k(X_i, X) + w_0$

$\therefore$ it can be represented as $g(X) = \sum_{i=1}^{N} \lambda_i z_i k(X_i, X) + w_0$

(d) We learned that for vectors that do not violate the margin[1] (i.e. $z_j(\mathbf{w}^T\mathbf{u}_j + w_0) - 1 > 0$), the Lagrange multiplier is zero, i.e. $\lambda_j = 0$. On the other hand, for vectors on the margin ($z_j(\mathbf{w}^T\mathbf{u}_j + w_0) - 1 = 0$), $\lambda_j \neq 0$. Show that, consequently, one can find a vector $\mathbf{x}_j$ for which $\lambda_j \neq 0$ and calculate w_0 as $w_0 = 1/z_j - \sum_{i=1}^{N} \lambda_i z_i \kappa(\mathbf{x}_i, \mathbf{x}_j)$.

(d) $\because z_j (w^T u_j + w_0) - 1 = 0$

$\therefore z_j (\sum_{i=1}^{N} \lambda_i z_i k(X_i, X_j) + w_0) = 1$

$\therefore \sum_{i=1}^{N} \lambda_i z_i k(X_i, X_j) + w_0 = z_j \quad (\because z_j^2 = 1)$

$\therefore w_0 = z_j - \sum_{i=1}^{N} \lambda_i z_i k(X_i, X_j)$

$\therefore w_0 = \frac{1}{z_j} - \sum_{i=1}^{N} \lambda_i z_i k(X_i, X_j) \quad (\because z_j = \pm 1,\ z_j = \frac{1}{z_j})$

$\therefore$ choose $X_2 = [1, 0]^T$, $z_2 = 1$

$w_0 = 1 - \lambda_1 z_1 k(X_1, X_2) - \lambda_2 z_2 k(X_2, X_2) - \lambda_3 z_3 k(X_3, X_2) - \lambda_4 z_4 k(X_4, X_2)$

$= 1 + \frac{8}{3} \times 1 - 1 \times 4 - \frac{2}{3} \times 1 - 0$

$= -1$

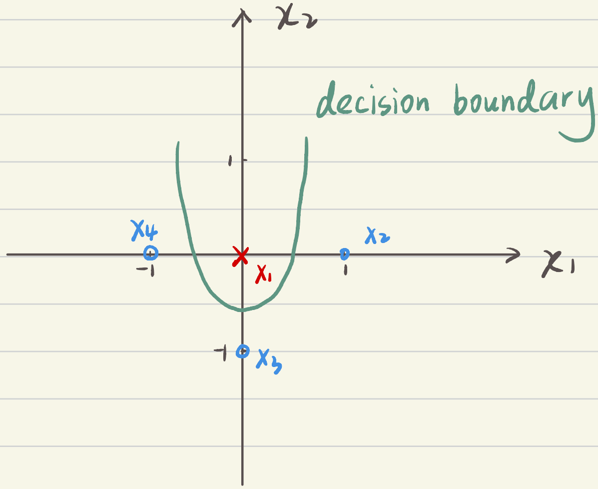
(e) Sketch the decision boundary for this data set based on parts (4c) and (4d).

(e) $\because g(X) = \sum_{i=1}^{4} \lambda_i z_i k(X_i, X) + w_0$

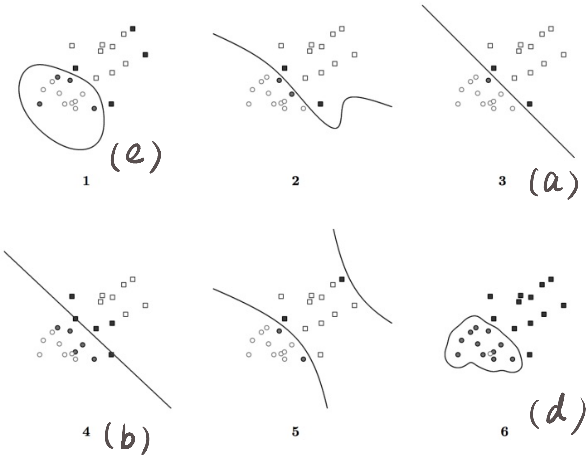
$= -\frac{8}{3} k(X_1, X) + k(X_2, X) + \frac{2}{3} k(X_3, X) + k(X_4, X) - 1$

$\therefore$ let $X = [x_1, x_2]^T$

$\therefore g(X) = -\frac{8}{3} + (x_1+1)^2 + \frac{2}{3}(-x_2+1)^2 + (-x_1+1)^2 - 1 = 0$



5. In the following figure, there are different SVMs with different decision boundaries. The training data is labeled as $z_i \in \{-1, 1\}$, represented as circles and squares respectively. Support vectors are drawn in solid circles. Determine which of the scenarios described below matches one of the 6 plots (note that one of the plots does not match any scenario). Each scenario should be matched to a unique plot. Explain your reason for matching each figure to each scenario. (10 pts)



(a) A soft-margin linear SVM with $C = 0.02$

(b) A soft-margin linear SVM with $C = 20$

(c) A hard-margin kernel SVM with $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j + (\mathbf{x}_i^T \mathbf{x}_j)^2$

(d) A hard-margin kernel SVM with $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp(-5\|\mathbf{x}_i - \mathbf{x}_j\|^2)$

(e) A hard-margin kernel SVM with $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\frac{1}{5}\|\mathbf{x}_i - \mathbf{x}_j\|^2)$

(a) → 3 : linear boundary with wider margin (small c)
(b) → 4 : linear boundary with tighter margin (large c)
(c) → 5 : smooth hyperbolic boundary, global quadratic curve
(d) → 6 : complex, tight boundary around circles
(e) → 1 : smother elliptical boundary