① $D_{w_1} = \{0, 2, 5\}$ $(h=1)$

$D_{w_2} = \{4, 7\}$

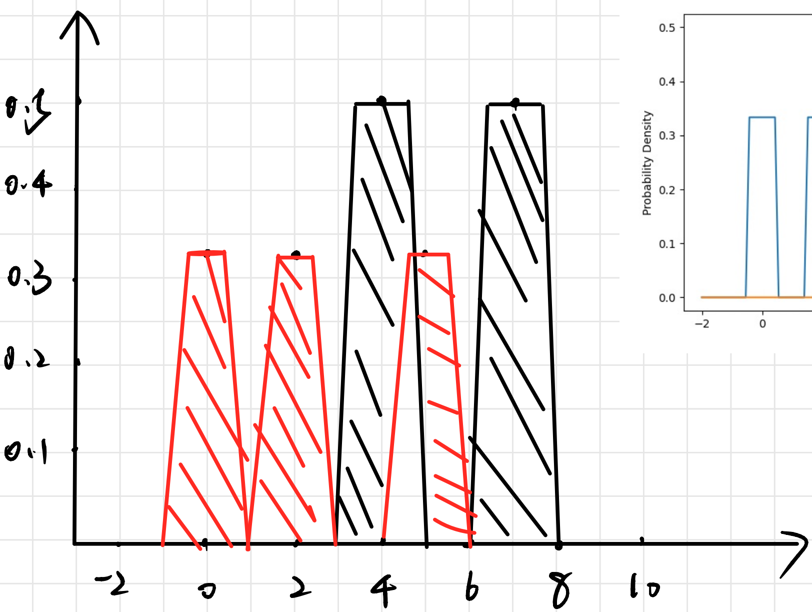
$\varphi(u) = \{1,\ |u| \le h/2\ \ 0,\ \text{else}\}$

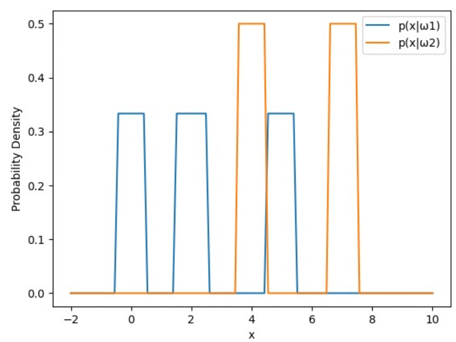
$$p(x|w_1) = \frac{1}{N_{w_1}} \cdot \Sigma \varphi\left(\frac{x-x_i}{h}\right) = \frac{1}{3} \cdot (\varphi(x-0) + \varphi(x-2) + \varphi(x-5))$$

$$= \frac{1}{3} \cdot (\varphi(x) + \varphi(x-2) + \varphi(x-5))$$

$$p(x|w_2) = \frac{1}{N_{w_2}} \cdot \Sigma \varphi\left(\frac{x-x_i}{h}\right) = \frac{1}{2}(\varphi(x-4) + \varphi(x-7))$$

$$= \frac{1}{2}(\varphi(x-4) + \varphi(x-7))$$





(b) $P(w_1) = \frac{3}{5}$ $P(w_2) = \frac{2}{5}$

(c) $P(w_1|x) = P(x|w_1)P(w_1) / P(x)$

$= P(x|w_1)P(w_1) / P(x|w_1)P(w_1) + P(x|w_2)P(w_2)$

$P(w_2|x) = P(x|w_2)P(w_2) / P(x)$

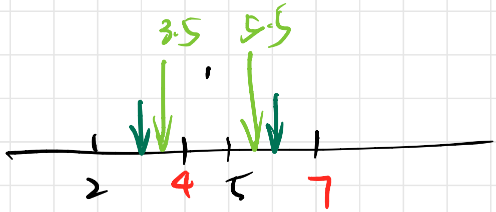
for the decision boundry, it happens when $P(w_1|x) = P(w_2|x)$

as ploted above, the decision region

and the decision boundry lies when $P(w_1|x) = P(w_2|x)$

② $D_{w_1} = \{2, 5\}$

$D_{w_2} = \{4, 7\}$



(a) -

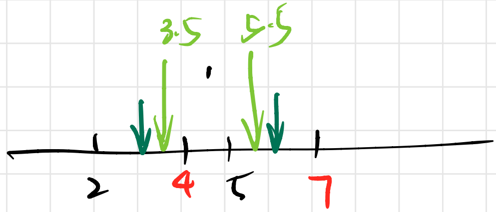
$(-\infty, 3.5)$ = nearest neighbors: $\{2, 4\}$ $\begin{cases} (-\infty, 3) : w_1 \\ (3, 3.5) : w_2 \end{cases}$

$(3.5, 5.5)$ = nearest neighbors: $\{4, 5\}$ $\begin{cases} (3.5, 4.5) : w_2 \\ (4.5, 5.5) : w_1 \end{cases}$

$(5.5, \infty)$: nearest neighbors: $\{5, 7\}$ $\begin{cases} (5.5, 6) : w_1 \\ (6, \infty) : w_2 \end{cases}$

b. $P(w_1) = \frac{2}{4} = \frac{1}{2}$

$P(w_2) = \frac{2}{4} = \frac{1}{2}$



c. $\begin{cases} P(w_1|x) > P(w_2|x) & w_1 \\ P(w_1|x) < P(w_2|x) & w_2 \\ P(w_1|x) = P(w_2|x) & \text{boundry} \end{cases}$

$(-\infty, 3.5)$ = nearest neighbors: $\{2, 4\}$ $\begin{cases} (-\infty, 3) = w_1 \\ (3, 3.5) = w_2 \end{cases}$

$(3.5, 5.5)$ = nearest neighbors: $\{4, 5\}$ $\begin{cases} (3.5, 4.5) = w_2 \\ (4.5, 5.5) = w_1 \end{cases}$

$(5.5, \infty)$ = nearest neighbors: $\{5, 7\}$ $\begin{cases} (5.5, 6) = w_1 \\ (6, \infty) = w_2 \end{cases}$

$x = 3.5 : w_2$; $x = 5.5 : w_2$

decision boundries $x = 3$, $x = 4.5$, $x = 6$

(d)

estimated Posterior probability for each region in the feature space:

In region, $(-\infty, 3.5)$, $(3.5, 5.5)$, $(5.5, \infty)$ we have $P(w_1|x) = P(w_2|x) = \frac{1}{2}$

however, in $(-10, 3)$ data set is prone to be defined as w_1,
since they are closer to 2.

$(3, 4.5)$ w_2

$(4.5, 6)$ w_1

$(6, 10)$ w_2

for highest Posterior Probability

$(-10, 3.5)$ $P(w_1|x) > P(w_2|x)$: w_1

$(3.5, 5.5)$ $P(w_1|x) = P(w_2|x)$ $\xRightarrow{\text{break tie}}$ w_2

$(5.5, 10)$ $P(w_1|x) < P(w_2|x)$ $= w_2$

both classifier would break the tie to support w_2

however, the classifier based on estimating $P(w_i|x)$ using the nearest neighbors provides more detailed information

3. $u = [\varphi_1(x), \varphi_2(x)]^T \quad \varphi_1(x) = x_2^2 - 2x_1 + 3 \quad \varphi_2(x) = x_1^2 - 2x_2 - 3$

$u_1 = [0-2+3, 1-3] = [1, -2] \quad -1 \quad (1, 0)$

$u_2 = [1-0+3, 0-2-3] = [4, -5] \quad -1 \quad (0, 1)$

$u_3 = [1-0+3, 0+2-3] = [4, -1] \quad -1 \quad (0, -1)$

$u_4 = [0+2+3, 1-0-3] = [5, -2] \quad 1 \quad (-1, 0)$

$u_5 = [4-0+3, 0-4-3] = [7, -7] \quad 1 \quad (0, 2)$

$u_6 = [4-0+3, 0+4-3] = [7, 1] \quad 1 \quad (0, -2)$

$u_7 = [0+4+3, 4-0-3] = [7, 1] \quad 1 \quad (-2, 0)$

$y = [-1, -1, -1, 1, 1, 1, 1]$

$P_1 = [1, 4, 4, 5, 7, 7, 7]$

$P_2 = [-2, -5, -1, -2, -7, 1, 1]$

$y = w_1\varphi_1(x) + w_2\varphi_2(x) + b$

$$\left\{\begin{array}{l} w_1 - 2w_2 + b \\ 4w_1 - 5w_2 + b \\ 4w_1 - w_2 + b \end{array}\right. = -1$$

$$\left\{\begin{array}{l} 5w_1 - 2w_2 + b \\ 7w_1 - 7w_2 + b \\ 7w_1 + w_2 + b \\ 7w_1 + w_2 + b \end{array}\right. = 1$$

$0 = 0.32P_1 + 0.02P_2 - 1.3$

is the equation of hyperplane

Code snippet

```
w1=1; w2=1; lr=0.01;
lms_error=1; bias=1;
while lms_error > 0.1
    b=[bias bias bias bias bias bias bias]
    act_y = [(b+w1*p1+w2*P2)>=0] -
        [(b+w1*P1+w2*P2)<=0];
    error = y - act_y;
    lms_error = 0.5*sum(abs(error));
    w1 = w1+lr*error*transpose(P1);
```

$w_2 = w_2 + lr * error * transpose(p_2)$;
bias = bias + lr * error * transpose(acty)

4. $K(x_i, x_j) = (x_i^T x_j + 1)^2$

(a) $K(x_1, x_1) = (x_1^T x_1 + 1)^2 = (0^T 0 + 1)^2 = 1$

$K(x_2, x_2) = (1+1)^2 = 4$

$K(x_3, x_3) = (1+1)^2 = 4$

$K(x_4, x_4) = (1+1)^2 = 4$

$K(x_1, x_2) = K(x_2, x_1) = 1$

$K(x_1, x_3) = K(x_3, x_1) = 1$

$K(x_1, x_4) = K(x_4, x_1) = 1$

$K(x_2, x_3) = K(x_3, x_2) = 1$

$K(x_2, x_4) = K(x_4, x_2) = 0$

$K(x_3, x_4) = K(x_4, x_3) = 1$

maximise =

$$L(\lambda) = \sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j z_i z_j (x_i^T x_j + 1)^2 \quad (\sum_{i=1}^N \lambda_i z_i = 0)$$

$$\Rightarrow \text{maximize: } \sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j z_i z_j K(x_i, x_j)$$

$$\begin{cases} w = \sum_{i=1}^N \lambda_i z_i x_i \\ b = z_k - \sum_{i=1}^N \lambda_i z_i K(x_i, x_k) \end{cases}$$

(k = index of any support vector with non-sero λ_i)

(b) $$\frac{\partial L(\lambda)}{\partial \lambda_i} = 1 - \sum_{j=1}^{N} \lambda_j z_i z_j k(x_i, x_j) - \sum a_j z_i = 0 \quad ①$$

$$\frac{\partial L}{\partial a} = \sum_{i=1}^{N} \lambda_i z_i = 0$$

substitute back to eliminate a

$$a_i = \frac{-\lambda_i z_i}{k(x_i, x_j)} \quad ②$$

substitute ② to ①

$$1 - \sum_{j=1}^{N} \lambda_j z_i z_j k(x_i, x_j) - \sum_{i=1}^{N} \left(\frac{-\lambda_j z_i}{k(x_i, x_j)} \right) z_i = 0$$

$$\lambda_i = \frac{1}{2 \cdot k(x_i, x_j) \cdot \sum_{j=1}^{N} \lambda_j z_i z_j k(x_i \cdot x_j)}$$

optimal λ_i = [0.74996703 0.25003297 0.25003297 0.25003297]

w = [0.74996703 0.25003297]

b = 0.4999022834131266

```
C:\Users\11190\Desktop\1\venv\Scripts\python.exe C:\Users\11190\Desktop\1\main.py
Optimal λ_i values: [0.74996703 0.25003297 0.25003297 0.25003297]
w: [0.74996703 0.25003297]
b: 0.4999022834131266

Process finished with exit code 0
```

decision boundary = $0.74996703 \cdot x_1 + 0.25003297\, x_2 + 0.4999022834131266$

(c) $w^T u + w_0 = 0$, $w = \sum_{i=1}^N \lambda_i z_i x_i$

substitute expression for w into decision boundry equation.

$$(\sum_{i=1}^N \lambda_i z_i x_i)^T u + w_0 = 0$$

$$\sum_{i=1}^N (\lambda_i z_i x_i)^T u + w_0 = 0 \qquad (a^T b = \langle a, b\rangle)$$

$$\sum_{i=1}^N \langle \lambda_i z_i x_i, u\rangle + w_0 = 0$$

since $\langle \lambda_i z_i x_i, u\rangle = \lambda_i z_i \langle x_i, u\rangle$

$$\therefore \sum_{i=1}^N \lambda_i z_i \langle x_i, u\rangle + w_0 = 0$$

$$\therefore \sum_{i=1}^N \lambda_i z_i k(x_i, u) + w_0 = 0$$

$\therefore$ for decison boundry $g(x) = \sum_{i=1}^N \lambda_i z_i k(x_i, x) + w_0$.

(d)

$$z_j (w^T x_j + w_0) - 1 = 0.$$

substitute

$$\Rightarrow z_j ((\sum_{i=1}^N \lambda_i z_i x_i)^T x_j + w_0) - 1 = 0$$

$$\Rightarrow z_j (\sum_{i=1}^N (\lambda_i z_i x_i)^T x_j + w_0) - 1 = 0$$

$$\Rightarrow z_j (\sum_{i=1}^N \lambda_i z_i k(x_i, x_j) + w_0) - 1 = 0$$

$$w_0 = \frac{1}{z_j} - \sum_{i=1}^N \lambda_i z_i k(x_i, x_j)$$

e. the decision boundary is:

$$0.74997x_1 + 0.25003x_2 + 0.49990 = 0$$

5 (a) plot 4 : it should be a linear decision boundry for the slackness includes more support vectors than Q2 due to the lower penalty on slackness

(b) plot 3 :

it should be a linear decision boundry for the slackness includes less support vectors than Q1 due to the higher penalty on slackness

(c) plot 5 : for $k(x_i, x_j) = x_i^T x_j + (x_i^T x_j)^2$. The curve is either eclipse or hyperbolic curve.

(d) plot 6

for $k(x_i, x_j) = \exp(-5\|x_i - x_j\|^2)$. When $\gamma = 5$, which is relatively large, so the kernal value is big, and the classification is hard with fewer supporting vectors.

(e) plot 1 : when γ is small the kernal values is actually small and the classification has more supporting vectors