Homework 9 - Part A

Exercise 1

$x_1 = (1, 10) - \omega_1$

$x_2 = (2, 7) - \omega_2$

$x_3 = (3, 6) - \omega_1$

$x_4 = (4, 8) - \omega_2$

$x_5 = (5, 9) - \omega_2$

Entropy is $I = -\left(P(\omega_1)\log_2 P(\omega_1) + P(\omega_2)\log_2 P(\omega_2)\right) = 0.971$

$P(\omega_1) = 2/5 = 0.4, \quad P(\omega_2) = 3/5 = 0.6$

Computation of the entropy reduction for the 1^{st} coordinate and value $x_1 \leq 3$.

$x_1 \leq 3$
- $N_Y = 3$
  - $\omega_1 : 2$
  - $\omega_2 : 1$
- $N_N = 2$
  - $\omega_1 : 0$
  - $\omega_2 : 2$

Thus, $PY(\omega_1) = 2/3 = 0.666$

$PY(\omega_2) = 1/3 = 0.333$

$$IY = -\left(PY(\omega_1)\log_2 PY(\omega_1) + PY(\omega_2)\log_2 PY(\omega_2)\right) =$$
$$= -\left(0.666 \log_2 0.666 + 0.333 \log_2 0.333\right) =$$
$$= 0.918$$

$PN(\omega_1) = 0/2 = 0$

$PN(\omega_2) = 2/2 = 1$

$$IN = -\left(PN(\omega_1)\log_2 PN(\omega_1) + PN(\omega_2)\log_2 PN(\omega_2)\right) =$$
$$= 0$$

$$\Delta I = I - 3/5\, IY - 2/5\, IN = 0.971 - 3/5 \cdot 0.918 - 2/5 \cdot 0 \Rightarrow$$

$$\Rightarrow \Delta I = 0.42$$

Exercise 2

$L(\theta,\theta_0,\lambda) = \frac{1}{2}\theta^T\theta - \sum_{i=1}^{\tilde{N}} \lambda_i [y_i(\theta^T x_i + \theta_0) - 1]$ (1) is the Lagrangian function for SVM problem

The KKT conditions state that the minimizer satisfies:

a. $\frac{\partial L}{\partial \theta} = 0 \Rightarrow \theta - \sum_{i=1}^{\tilde{N}} \lambda_i y_i x_i = 0 \Rightarrow \theta = \sum_{i=1}^{\tilde{N}} \lambda_i y_i x_i$ (2)

b. $\frac{\partial L}{\partial \theta_0} = 0 \Rightarrow \sum_{i=1}^{\tilde{N}} \lambda_i y_i = 0$ (3)

Now, let's replace (2), (3) to (1):

$$L = \frac{1}{2}\left(\sum_{i=1}^{\tilde{N}} \lambda_i y_i x_i\right)^T \sum_{i=1}^{\tilde{N}} \lambda_i y_i x_i - \sum_{i=1}^{\tilde{N}} \lambda_i \left[y_i \left(\left(\sum_{j=1}^{\tilde{N}} \lambda_j y_j x_j \right)^T x_i + \theta_0 \right) - 1 \right] =$$

$$= \frac{1}{2} \sum_{i=1}^{\tilde{N}} \lambda_i y_i x_i^T \sum_{i=1}^{\tilde{N}} \lambda_i y_i x_i - \left[\sum_{i=1}^{\tilde{N}} \lambda_i y_i \sum_{j=1}^{\tilde{N}} \lambda_j y_j x_j^T x_i + \cancel{\lambda_i y_i} \sum_{i=1}^{\tilde{N}} \lambda_i y_i \theta_0 - \sum_{i=1}^{\tilde{N}} \lambda_i \right] =$$

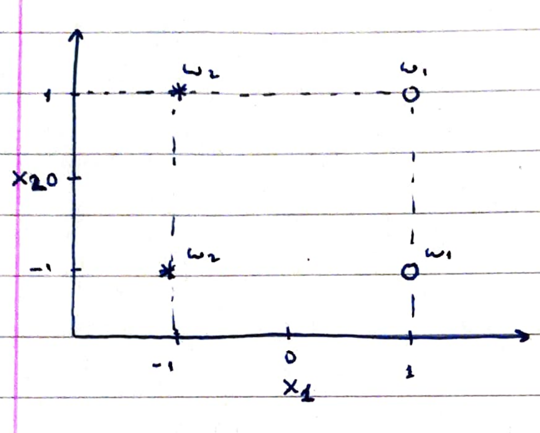
$$= \frac{1}{2} \sum_{i=1}^{\tilde{N}} \sum_{j=1}^{\tilde{N}} \lambda_i \lambda_j y_i y_j x_i^T x_j - \sum_{i=1}^{\tilde{N}} \sum_{j=1}^{\tilde{N}} \lambda_i \lambda_j y_i y_j x_i^T x_j + \sum_{i=1}^{\tilde{N}} \lambda_i \Rightarrow$$

$\Rightarrow L(\lambda_1, \lambda_2 \ldots \lambda_N) = \sum_{i=1}^{\tilde{N}} \lambda_i - \frac{1}{2} \sum_{i=1}^{\tilde{N}} \sum_{j=1}^{\tilde{N}} \lambda_i \lambda_j y_i y_j x_i^T x_j$: Dual Wolfe representation

Max $L(\lambda)$ subject to $\sum_{i=1}^{\tilde{N}} \lambda_i y_i = 0$, $\lambda \geq 0$

Exercise 3

(a)



- diagram also in pdf -

Classes are linearly seperable by $x_1 = 0$ or $x_2 = 0$

(b) For the Lagrangian function : $N = 4$, $y_1 = 1$, $y_2 = 1$, $y_3 = -1$, $y_4 = -1$

$$L(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \sum_{i=1}^{4} \lambda_i - \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \lambda_i \lambda_j y_i y_j x_i^T x_j$$

- $y_1 y_1 x_1^T x_1 = 1 \cdot 1 \; [-1, 1] \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 2$
- $y_1 y_2 x_1^T x_2 = 0$
- $y_1 y_3 x_1^T x_3 = 2$
- $y_1 y_4 x_1^T x_4 = 0$
- $y_2 y_2 x_2^T x_2 = 2$
- $y_2 y_3 x_2^T x_3 = 0$
- $y_2 y_4 x_2^T x_4 = 2$
- $y_3 y_3 x_3^T x_3 = 2$
- $y_3 y_4 x_3^T x_4 = 0$
- $y_4 y_4 x_4^T x_4 = 2$

$$L(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 - \frac{1}{2}(2\lambda_1^2 + 2\lambda_1\lambda_3 + 2\lambda_2^2 + 2\lambda_2\lambda_4 + 2\lambda_3\lambda_1 + $$
$$+ 2\lambda_3^2 + 2\lambda_4\lambda_2 + 2\lambda_4^2 \Rightarrow$$

$$\Rightarrow L(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 - \lambda_1^2 - \lambda_2^2 - \lambda_3^2 - \lambda_4^2 - 2\lambda_1\lambda_3 - 2\lambda_2$$

$$\frac{\partial L}{\partial \lambda_1} = 1 - 2\lambda_1 - 2\lambda_3 = 0$$

$$\frac{\partial L}{\partial \lambda_2} = 1 - 2\lambda_2 - 2\lambda_4 = 0$$

$$\frac{\partial L}{\partial \lambda_3} = 1 - 2\lambda_3 - 2\lambda_1 = 0$$

$$\frac{\partial L}{\partial \lambda_4} = 1 - 2\lambda_2 - 2\lambda_4 = 0$$

$\Rightarrow$

$$\lambda_1 + \lambda_3 = \frac{1}{2}$$

$$\lambda_2 + \lambda_4 = \frac{1}{2}$$

- Since Lagrangian function runs under the constraint $\sum_{i=1}^{\tilde{}} \lambda_i y_i = 0$ we have that : $\lambda_1+\lambda_2-\lambda_3-\lambda_4=0$

$$\left.\begin{array}{l} \lambda_1+\lambda_3=\frac{1}{2} \\ \lambda_2+\lambda_4=\frac{1}{2} \\ \lambda_1+\lambda_2-\lambda_3-\lambda_4=0 \\ \quad \lambda_i \geq 0 \end{array}\right\} \Rightarrow \begin{array}{l} \lambda_1=\lambda_4 \\ \lambda_2=\lambda_3 \\ \lambda_2=\frac{1}{2}-\lambda_1 \\ \lambda_3=\frac{1}{2}-\lambda_4 \end{array}$$

Thus $\lambda_1=\lambda_4=a$ and $\lambda_2=\lambda_3=\frac{1}{2}-a$, $a\in[0,\frac{1}{2}]$

- Since Lagrangian function runs under the constraint $\theta=\sum_{i=1}^{\tilde{}} \lambda_i y_i x_i$ we have that:

$$\theta = \lambda_1 \cdot y_1 x_1 + \lambda_2 y_2 x_2 + \lambda_3 y_3 x_3 + \lambda_4 y_3 x_4 = \lambda_1 x_1 + \lambda_2 x_2 - \lambda_3 x_3 - \lambda_4 y_4 =$$

$$= \lambda_1 \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} -1 \\ -1 \end{bmatrix} + \lambda_3 \begin{bmatrix} 1 \\ -1 \end{bmatrix} - \lambda_4 \begin{bmatrix} 1 \\ 1 \end{bmatrix} =$$

$$\cancel{= a\begin{bmatrix} -1 \\ 1 \end{bmatrix} + (\tfrac{1}{2}-a)\begin{bmatrix} -1 \\ -1 \end{bmatrix} - (\tfrac{1}{2}-a)\begin{bmatrix} 1 \\ -1 \end{bmatrix} - a\begin{bmatrix} 1 \\ 1 \end{bmatrix} =}$$

$$\cancel{= \begin{bmatrix} -a \\ a \end{bmatrix} + \begin{bmatrix} -\frac{1}{2}+a \\ -\frac{1}{2}+a \end{bmatrix} + \begin{bmatrix} -\frac{1}{2}+a \\ \frac{1}{2}-a \end{bmatrix} + \begin{bmatrix} -a \\ -a \end{bmatrix} =}$$

$$\cancel{= \begin{bmatrix} -a-\frac{1}{2}+a-\frac{1}{2}+a+a \\ a-\frac{1}{2}+a+\frac{1}{2}-a+a \end{bmatrix}}$$

$$= \begin{bmatrix} -\lambda_1 \\ \lambda_1 \end{bmatrix} + \begin{bmatrix} -\lambda_2 \\ -\lambda_2 \end{bmatrix} + \begin{bmatrix} -\lambda_2 \\ \lambda_2 \end{bmatrix} + \begin{bmatrix} -\lambda_1 \\ -\lambda_1 \end{bmatrix} = \begin{bmatrix} -\lambda_1-\lambda_2-\lambda_2-\lambda_1 \\ \lambda_1-\lambda_2+\lambda_2-\lambda_1 \end{bmatrix} =$$

$$= \begin{bmatrix} -2\lambda_1-2\lambda_2 \\ 0 \end{bmatrix} = \begin{bmatrix} -2(\lambda_1+\lambda_2) \\ 0 \end{bmatrix} = \begin{bmatrix} -2\frac{1}{2} \\ 0 \end{bmatrix} \Rightarrow$$

$$\Rightarrow \theta = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

- for θ_0 we have that $\lambda_i[y_i(\theta^T x_i+\theta_0)-1]=0$, $i=1\dots4$

  Thus $\sum\lambda_i=0 \Rightarrow a=0$ or $a=\frac{1}{2}$

  or

  $\sum_{i=1}^{4} y_i(\theta^T x_i+\theta_0)-1=0 \Rightarrow$

$$\left.\begin{aligned}
&\Rightarrow y_1\begin{bmatrix}-1\\0\end{bmatrix}^T x_1+\theta_0-1=0 \Rightarrow [-1\ \ 0]\begin{bmatrix}-1\\1\end{bmatrix}+\theta_0-1=0\\
\text{and} &\Rightarrow y_2\begin{bmatrix}-1\\0\end{bmatrix}^T x_2+\theta_0-1=0 \Rightarrow [-1\ \ 0]\begin{bmatrix}-1\\-1\end{bmatrix}+\theta_0-1=0\\
\text{and} &\Rightarrow y_3\begin{bmatrix}-1\\0\end{bmatrix}^T x_3+\theta_0-1=0 \Rightarrow -[-1\ \ 0]\begin{bmatrix}1\\-1\end{bmatrix}+\theta_0-1=0\\
\text{and} &\Rightarrow y_4\begin{bmatrix}-1\\0\end{bmatrix}^T x_4+\theta_0-1=0 \Rightarrow -[-1\ \ 0]\begin{bmatrix}1\\1\end{bmatrix}+\theta_0-1=0
\end{aligned}\right\} \Rightarrow \theta_0=0$$

(c) In conclusion $\theta=[-1\ \ 0]^T$ and $\theta_0=0$. This is valid for any combination of λ_i's $(i=1\dots4)$, $\lambda_i>0$.