Mouselinos Spyridon
Homework 9.a
March 2020

Exersize 1

We have a two-class case with 5 points

Class ω_1

$$\left.\begin{array}{l} x_1 = (1,10) \\ x_3 = (3,6) \end{array}\right\} N_1 = 2$$

Class ω_2

$$\left.\begin{array}{l} x_2 = (2,7) \\ x_4 = (4,8) \\ x_5 = (5,9) \end{array}\right\} N_2 = 3$$

$N = N_1 + N_2 = 5$

Initially we have that the total entropy is

$$I(t) = -\sum_{i=1}^{M} P(\omega_i|t) \log_2 P(\omega_i|t)$$

where
- $P(\omega_1|t) = \dfrac{N_1}{N} = \dfrac{2}{5}$
- $\log_2 P(\omega_1|t) = \log_2 \dfrac{2}{5} = -1{,}3219$
- $P(\omega_2|t) = \dfrac{N_2}{N} = \dfrac{3}{5}$
- $\log_2 P(\omega_2|t) = \log_2 3/5 = -0{,}7369.$

So $$I(t) = -\sum_{i=1}^{2} P(\omega_i|t) \log_2 P(\omega_i|t) = 0{,}971$$

If we choose to use the split-node with the criterion $x_1 \leq 1$ we get two sets

$T_Y = \{ x_1 \}$ $N_{TY} = 1$

$N = N_{TY} + N_{TN} = 5$

$T_N = \{ x_2, x_3, x_4, x_5 \}$ $N_{TN} = 4$

So we calculate the entropy drop from the yes branch:

- $-\frac{N_{TY}}{N} I(t_Y) = -\frac{1}{5}\left(-\sum_{i=1}^{2} P(\omega_i|t_Y)\log_2 P(\omega_i|t_Y)\right)$

- Where $P(\omega_1|t_Y) = \frac{1}{1}$ and $\log_2 P(\omega_1|t_Y) = 0$

  $P(\omega_2|t_Y) = 0/1$ and $\log_2 P(\omega_2|t_Y) = -\infty$

So $-\frac{N_{TY}}{N} I(t_Y) = +\frac{1}{5}\left[1 \cdot 0 + 0 \cdot (-\infty)\right] = \frac{1}{5} \cdot 0 = 0$

- $-\frac{N_{TN}}{N} I(t_N) = -\frac{4}{5}\left(-\sum_{i=1}^{2} P(\omega_i|t_N)\log_2 P(\omega_i|t_N)\right)$

- Where $P(\omega_1|t_N) = 1/4$ $\log_2 P(\omega_1|t_N) = -2$

  $P(\omega_2|t_N) = 3/4$ $\log_2 P(\omega_2|t_N) = -0,4150$

$+\frac{4}{5}\left[1/4(-2) + 3/4(-0,4150)\right] = 0,64904$

Finally: $\Delta I = I - \frac{N_{TY}}{N} I(t_Y) - \frac{N_{TN}}{N} I(t_N) = 0,971 - 0 - 0,64904$
$= 0,3219$

Exersize 2

Let's start by taking the Lagrangian function of the SVM problem.

- $L(\theta, \theta_0, \lambda) = \frac{1}{2}\theta^T\theta - \sum_{i=1}^{N} \lambda_i [y_i(\theta^T x_i + \theta_0) - 1]$

According to the KKT conditions:

- $\frac{\partial L}{\partial \theta} = 0$

- $\frac{\partial L}{\partial \theta_0} = 0$

So: $\frac{\partial L}{\partial \theta} = 0 \Rightarrow \theta = \sum_{i=1}^{N} \lambda_i y_i x_i$

$\frac{\partial L}{\partial \theta_0} = 0 \Rightarrow 0 = \sum_{i=1}^{N} \lambda_i y_i$

Then we can replace them to the Lagrangian above

$$L(\theta, \theta_0, \lambda) = \frac{1}{2}(\Sigma \lambda_i x_j y_i)^T (\Sigma \lambda_i x_i y_i) - \Sigma \lambda_i [y_i ((\Sigma \lambda_i x_i y_i)^T x_i + \theta_0) - 1]$$

Thus we have:

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

$$\Rightarrow L(\theta, \theta_0, \lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j$$

The dual SVM problem is:

$$\max L(\theta, \theta_0, \lambda) \quad \text{wrt } \lambda\text{'s}$$

$$\text{subject to} \quad \sum_{i=1}^{N} (\lambda_i y_i) = 0$$

$$\lambda_i \geq 0 \ , i = 1, \dots, N$$

This is equivalent to the Wolfe dual representatio

$$\max_{\lambda \geq 0} L(\theta, \lambda)$$

$$\text{subject to} \quad \frac{\partial L(\theta, \lambda)}{\partial \theta} = 0$$

Exersize 3

(b) The two classes are linearly separable by any hyperplane of the form $x_1 = c$ with $c \in (-1, 1)$. However, because we demand the largest margin the most suitable hyperplane would be $x_1 = 0$.

In order to solve this problem we start by taking the dual Wolfe representation.

- $$J_1^*(\lambda) = \sum \lambda_i - \frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j x_i^T x_j$$

We have that $N = 4$, $y_1 = 1$, $y_2 = 1$, $y_3 = -1$, $y_4 = -1$

and that: $\begin{cases} x_1^T x_1 = 2 & x_1^T x_3 = -2 \\ x_1^T x_2 = 0 & x_1^T x_4 = 0 \end{cases}$

$\begin{cases} x_2^T x_1 = 0 & x_2^T x_3 = 0 \\ x_2^T x_2 = 2 & x_2^T x_4 = -2 \end{cases}$

$\begin{cases} x_3^T x_1 = -2 & x_3^T x_3 = 2 \\ x_3^T x_2 = 0 & x_3^T x_4 = 0 \end{cases}$

$\begin{cases} x_4^T x_1 = 0 & x_4^T x_3 = 0 \\ x_4^T x_2 = -2 & x_4^T x_4 = 2. \end{cases}$

After calculating all theese we plug them into the $J_1(\lambda)$ and we get:

$$J_1(\lambda) = L = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 - \lambda_1^2 - \lambda_2^2 - \lambda_3^2 - \lambda_4^2 - 2\lambda_1\lambda_3 - 2\lambda_2\lambda_4$$

Howerer this cost can be optimized wrt all different λ: $\frac{\partial L}{\partial \lambda_i} = 0$

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

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

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

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

We get the system

$$\begin{cases} 1 - 2\lambda_1 - 2\lambda_3 = 0 \\ 1 - 2\lambda_2 - 2\lambda_4 = 0 \\ 1 - 2\lambda_1 - 2\lambda_3 = 0 \\ 1 - 2\lambda_2 - 2\lambda_4 = 0 \end{cases} \Rightarrow \begin{cases} 1 - 2\lambda_1 - 2\lambda_3 = 0 \\ 1 - 2\lambda_2 - 2\lambda_4 = 0 \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} 2\lambda_1 + 2\lambda_3 = 2\lambda_2 + 2\lambda_4 \\ \lambda_1 + \lambda_3 = 1/2 \\ \lambda_2 + \lambda_4 = 1/2 \end{cases} \Rightarrow \begin{cases} \lambda_1 + \lambda_3 = 1/2 \\ \lambda_2 + \lambda_4 = 1/2 \end{cases}$$

Exersize 3 (cont)

We know so far $\lambda_1 + \lambda_3 = \frac{1}{2}$
$\lambda_2 + \lambda_4 = \frac{1}{2}$ $\Rightarrow \lambda_1 + \lambda_3 = \lambda_2 + \lambda_4$ (1)

and from the constraint $\sum \lambda_i y_i = 0 \Rightarrow$

$$\lambda_1 + \lambda_2 = \lambda_3 + \lambda_4 \quad (2)$$

From (1) and (2) we conclude that:

$$\lambda_1 = \lambda_4 \quad \text{and} \quad \lambda_2 = \lambda_3$$

and $0 \leq \lambda_1 \leq 1/2$, $0 \leq \lambda_2 \leq 1/2$ $\Rightarrow$ $0 \leq \lambda_3 \leq 1/2$, $0 \leq \lambda_4 \leq 1/2$

We can set $\lambda_1 = \lambda_4 = z$ and $\lambda_2 = \lambda_3 = 1/2 - z$

Now we can compute θ as follows:

$$\theta = \sum_{i=1}^{4} \lambda_i y_i x_i = z y_1 x_1 + (1/2 - z) y_2 x_2 + (1/2 - z) y_3 x_3 + z y_4 x_4$$

$$\theta = z[-1,1]^T + (1/2 - z)[-1,-1]^T - (1/2 - z)[1,-1]^T - z[1,1]^T = [-1,0]^T$$

θ_0 can be computed from the system:

$$\lambda_i [y_i(\theta^T x_i + \theta_0) - 1] = 0 \quad \forall i$$

$$\Rightarrow \begin{cases} -\theta_1 + \theta_2 + \theta_0 - 1 = 0 \\ -\theta_1 - \theta_2 + \theta_0 - 1 = 0 \\ -\theta_1 + \theta_2 - \theta_0 - 1 = 0 \\ -\theta_1 - \theta_2 - \theta_0 - 1 = 0 \end{cases} \Rightarrow$$

with $\theta = [\theta_1, \theta_2]^T = [-1, 0]^T$

lets substitute again

$$\Rightarrow \begin{cases} 1 + \theta_0 - 1 = 0 \\ 1 + \theta_0 - 1 = 0 \\ 1 - \theta_0 1 = 0 \\ 1 - \theta_0 1 = 0 \end{cases} \Rightarrow \begin{cases} \theta_0 = 0. \end{cases}$$

so the solution is $\theta = [-1, 0]$ $\theta_0 = 0$

Hyperplane is

$$g(x) = 0 \Rightarrow \begin{bmatrix} -1 \\ 0 \end{bmatrix} [x_1 \; x_2] + [0] = 0 \Rightarrow x_1 = 0$$

(c) Finally we know that regardless of the choice of λ's, as long as they satisfy that

$$\lambda_1 = \lambda_4 \quad \text{and} \quad \lambda_2 = \lambda_3 \quad \text{and}$$

that $0 \leq \lambda_1 = \lambda_4 \leq 1/2$ $\quad 0 \leq \lambda_2 = \lambda_3 \leq 1/2$

The parametric solution setting $\lambda_1 = u$ and $\lambda_2 = 1/2 - u$ will always give the same solution for θ