Yucheng Yang 2333448896

HW 3

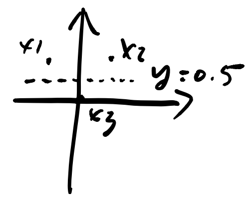
Problem 1:

(1.1) can't perfectly separated.

when the classifier $x \in \begin{cases} -1 \le x \le 0 & x_2, x_3 \text{ can't be separated} \\ x < -1 \text{ or } x > 1 & \text{all three can't} \\ 0 < x \le 1 & x_1, x_2 \text{ can't be separated.} \end{cases}$

(1.2)

$x_1 \to (-1, 1)$ $x_2 \to (1,1)$ $x_3 \to (0,0)$



According to the theorem, three points can be separated by a line like $y=0.5$ in $R2$.

(1.3)

$$\text{matrix } k = \begin{Bmatrix} \phi(x_1)\phi(x_1), \phi(x_1)\phi(x_2), \phi(x_1)\phi(x_3) \\ \phi(x_2)\phi(x_1), \phi(x_2)\phi(x_2), \phi(x_2)\phi(x_3) \\ \phi(x_3)\phi(x_1), \phi(x_3)\phi(x_2), \phi(x_3)\phi(x_3) \end{Bmatrix}$$

$$= \begin{Bmatrix} x_1^2+x_1^4 & x_1x_2+x_1^2x_2^2 & x_1x_3+x_1^2x_3^2 \\ x_1x_2+x_1^2x_2^2 & x_2^2+x_2^4 & x_2x_3+x_2^2x_3^2 \\ x_3x_1+x_3^2x_1^2 & x_3x_2+x_3^2x_2^2 & x_3^2+x_3^4 \end{Bmatrix}$$

(1.4)

We ignore slack variables because it's separable

primal: $\min\limits_{k,b} \frac{1}{2}\|w\|^2$ s.t $y_i(wx_i+b) \ge 1 \quad i=1,2,3$

$\therefore \quad L(w,b,a) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{3} a_i (y_i (w x_i + b) - 1) =$

$$\min_{w,b} \max_{a_i \geq 0} L(w,b,a)$$

dual: $\max_{a_i \geq 0} \min_{w,b} L(w,b,a)$.

(1.5) According to KKT, solving dual formular equals solving primal formular.

let $\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{3} a_i y_i x_i$

let $\frac{\partial L}{\partial b} = 0 \Rightarrow 0 = \sum_{i=1}^{3} a_i y_i$

$$L(w,b,a) = \frac{1}{2} w^T w - w^T \sum_{i=1}^{n} a_i y_i x_i - b \sum_{i=1}^{n} a_i y_i + \sum_{i=1}^{n} a_i$$

$$= \sum_{i=1}^{3} a_i - \frac{1}{2} \sum_{i=1,j=1}^{3} a_i a_j y_i y_j x_i x_j$$

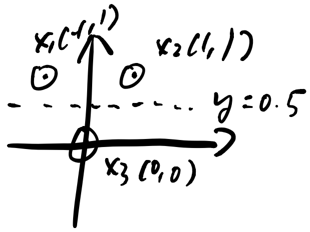
$\therefore \quad \min_{a} \frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} (a_i a_j y_i y_j x_i x_j) - \sum_{i=1}^{3} a_i \quad s.t. \sum_{i=1}^{3} a_i y_i = 0$

$a_i \geq 0 \quad i = 1,2,3$

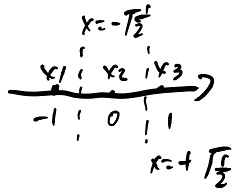
$w^* = \sum_{i=1}^{3} a_i^* y_i x_i = (0, 1) \quad b^* = y_i - \sum_{j=1}^{N} a_j^* y_j x_i x_j = -0.5$

$\therefore \quad y = 0.5$

(1.6)



x_1, x_2, x_3 are SV



$w\phi(x) + b = 0 \rightarrow (0,1) \cdot (x, x^2) - 0.5 = 0$

$x^2 - 0.5 = 0$

$x = \pm\sqrt{\frac{1}{2}}$

problem 2

2.1 entropy:

$T_1 A$: $I_H(D_P) = -(0.5\log_2(0.5) + 0.5\log_2(0.5)) = 1$

$I_H(D_{left}) = -(\frac{50}{200}\log_2\frac{50}{200} + \frac{150}{200}\log_2\frac{150}{200}) = 0.81$

$I_H(D_{right}) = -(\frac{150}{200}\log_2\frac{150}{200} + \frac{50}{200}\log_2\frac{50}{200}) = 0.81$

A: $IG_H = 1 - \frac{1}{2}\times 0.81 - \frac{1}{2}\times 0.81 = 0.19$

$T_2 B$:

$I_H(D_p) = 1$

$I_H(D_{left}) = 0$

$I_H(D_{right}) = -(\frac{100}{300}\log_2\frac{100}{300} + \frac{200}{300}\log_2\frac{200}{300}) = 0.92$

$IG_H = 1 - \frac{300}{400}\cdot 0.92 - 0 = 0.31$

weighted Gini

$T_1 A$: $I_G(D_P) = 1 - (0.5^2 + 0.5^2) = 0.5$

$I_G(D_{left}) = 1 - ((\frac{1}{4})^2 + (\frac{3}{4})^2) = \frac{3}{8}$

$I_G(D_{right}) = 1 - ((\frac{3}{4})^2 + (\frac{1}{4})^2) = \frac{3}{8}$

$I_G = 0.5 - \frac{1}{2}\cdot\frac{3}{8} - \frac{1}{2}\times\frac{3}{8} = 0.125$

$T_2 B$:

$I_G(D_p) = 1 - (0.5^2 + 0.5^2) = 0.5$

$I_G(left) = 0$

$I_G(right) = 1 - ((\frac{100}{300})^2 + (\frac{200}{300})^2) = \frac{4}{9}$

$I_G = 0.5 - \frac{3}{4}\cdot\frac{4}{9} - 0 = \frac{1}{2} - \frac{3}{9} = \frac{1}{6}$

classification error:

T_1 A. $I_E(D_p) = 1-0.5 = 0.5$

$I_E(D\ left) = 1 - \frac{150}{200} = 0.25$

$I_E(D\ right) = 1 - \frac{150}{200} = 0.25$

$IG_E = 0.5 - \frac{1}{2} \times 0.25 - \frac{1}{2} \times 0.25 = 0.25$

T_2 B. $I_E(D_p) = 0.5$

$I_E(D_{left}) = 1-1 = 0$ $\quad I_E(D\ right) = 1 - \frac{200}{300} = \frac{1}{3}$

$IG_E = 0.5 - \frac{300}{400} \times \frac{1}{3} = 0.25$

2.2

we see that the entropy $T_1 < T_2$ means T_2 is more difficult and complicated.
weighted Gini $T_1 < T_2$ also means T_1 is better than T_2.
And we know that $E_1 = E_2$ ~~means that~~ they did same in this data set. However, T_1 is obviously better than T_2. So that we can't use E. We should use Gini or entropy.

problem 3

(3.1) Because it is convex, so

let $f(\beta_t)' = \varepsilon_t(e^{\beta t} + e^{-\beta t}) - e^{-\beta t} = 0$

$\therefore \frac{1-\varepsilon_t}{\varepsilon_t} = e^{2\beta t}$

$\therefore \beta_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) = \beta_t^*$

(3, 2)

$$\varepsilon = \sum_{n: h_t(x_n) \neq y_n} D_t(n)$$

$$\underset{n: h_t(x_n) \neq y_n}{D_{t+1}(n)} = \frac{D_t(n) \cdot e^{a_t}}{Z_n} \quad a_t = \frac{1}{2}\left(\ln \frac{1-\varepsilon}{\varepsilon}\right)$$

$$\therefore \sum_{n: h_t(x_n) \neq y_n} D_{t+1}(n) = \frac{\sum_{n: h_t(x_n) \neq y_n} D_t(n) \cdot e^{a_t}}{\sum_{n: h_t(x_n) \neq y_n} D_t(n) \cdot e^{a_t} + \sum_{n: h_t(x_n) = y_n} D_t(n) \cdot e^{-a_t}}$$

$$= \frac{\varepsilon e^{a_t}}{\varepsilon e^{a_t} + (1-\varepsilon) e^{-a_t}} = \frac{1}{1 + \frac{(1-\varepsilon)}{\varepsilon} e^{-2a_t}} = \frac{1}{1+1} = \frac{1}{2}$$