1)

a,

Iterations	E_{in}	binary classification error on training data	binary classification error on testing data	Time(s)
10^4	0.5847	0.3092	0.3172	16.4389
10^5	0.4937	0.2237	0.2068	164.3292
10^6	0.4354	0.1513	0.1310	1659.7216

↳ Increasing the maximum number of iterations will decrease the cross-entropy error (E_{in}) as well as the binary classification error on both training and testing data.

b)

Learning rate η	E_{in}	binary classification on training dataset	binary classification on testing dataset	Iterations required to terminate	Time (s)
0.01	0.40742	0.17105	0.51034	6738	11.8585
0.1	0.40742	0.17105	0.51034	673	1.1897
1	0.407416	0.17105	0.51034	67	0.1206
4	0.407403	0.17105	0.51034	16	0.0281
7	0.407387	0.17105	0.51034	10	0.0200
7.5	0.407389	0.17105	0.51034	22	0.0400
7.6	0.407389	0.17105	0.51034	37	0.0662
7.7	0.40739	0.17105	0.51034	217	0.3685

↳ Normalizing the data increases the performance of the model by reducing the time the training process took.

Changing the learning rate η may affect the number of iterations required to terminate.

2, LFD Problem 2.22

$$E_{out}(g^{(D)}) = E_{x,y}[(g^{(D)}(x) - y(x))^2]$$

$$\hookrightarrow ① \ E_D[E_{out}(g^{(D)})] = E_D[E_{x,y}[(g^{(D)}(x) - y(x))^2]]$$

$$= E_{x,y}[E_D[(g^{(D)}(x) - y(x))^2]]$$

$$= E_{x,y}[E_D[g^{(D)}(x)^2] - 2\bar{g}(x)y(x) + y(x)^2]$$

$$= E_{x,y}[E_D[g^{(D)}(x)^2] - 2\bar{g}(x)(f(x)+\varepsilon) + (f(x)+\varepsilon)^2]$$

$$= E_{x,y}[E_D[g^{(D)}(x)^2] - \bar{g}(x)^2 + \bar{g}(x)^2 - 2\bar{g}(x)f(x) + f(x)^2 + \varepsilon^2$$
$$- 2\bar{g}(x)\varepsilon + 2f(x)\varepsilon]$$

$$= E_{x,y}[\underbrace{(E_D[g^{(D)}(x)^2] - \bar{g}(x)^2)}_{\shortparallel} + \underbrace{(\bar{g}(x) - f(x))^2}_{bias(x)} + \varepsilon^2 - 2\varepsilon(\bar{g}(x) - f(x))]$$

$$E_D[g^{(D)}(x)^2] - 2E_D[g^{(D)}(x)]\bar{g}(x) + \bar{g}(x)^2$$

$$= E_D[g^{(D)}(x)^2 - 2g^{(D)}(x)\bar{g}(x) + \bar{g}(x)^2]$$

$$= E_D[(g^{(D)}(x) - \bar{g}(x))^2] = var(x)$$

$$① = E_{x,y}[var(x)] + E_{x,y}[bias(x)] + E_{x,y}[\varepsilon^2] - 2E_{x,y}[(\bar{g}(x) - f(x))\varepsilon]$$

$$= var + bias + \underbrace{E_{x,y}[\varepsilon^2] - 2E_{x,y}[(\bar{g}(x) - f(x))\varepsilon]}_{Ⓐ}$$

$$A = E_x\left[E_\varepsilon\left[\varepsilon^2\right]\right] - 2E_x\left[\left(\bar{g}(x) - f(x)\right)E_\varepsilon[\varepsilon]\right]$$

$$= \sigma^2 + 0$$

$$\Rightarrow ① = E_D\left[E_{out}\left(g^{(D)}\right)\right] = \text{var} + \text{bias} + \sigma^2$$

3. LFD Problem 2.24

a, $\begin{cases} x_1^2 = ax_1 + b \\ x_2^2 = ax_2 + b \end{cases} \rightarrow \begin{cases} a = x_1 + x_2 \\ b = -x_1 x_2 \end{cases}$

$\bar{g}(x) = \bar{a}x + b$, where $\begin{cases} \bar{a} = E[a] = E[x_1 + x_2] \\ \bar{b} = E[b] = E[-x_1 x_2] \end{cases}$

$\bar{a} = E[x_1] + E[x_2] = 0$

$\bar{b} = E[-x_1 x_2] = \frac{-1}{4} \int_{-1}^{1} \int_{-1}^{1} (x_1 x_2) dx_1 dx_2 = \frac{-1}{4} \cdot 0 = 0$

$\therefore \bar{g}(x) = 0$

b,

- $\bar{g}(x)$: generate dataset D n times & compute

$g(x) = ax + b \rightarrow$ then $\bar{g}(x) = \bar{a}x + \bar{b}$

where $\bar{a} = \frac{1}{N} \sum_{i=1}^{N} (x_1 + x_2)$ & $\bar{b} = \frac{-1}{N} \sum_{i=1}^{N} (x_1 x_2)$

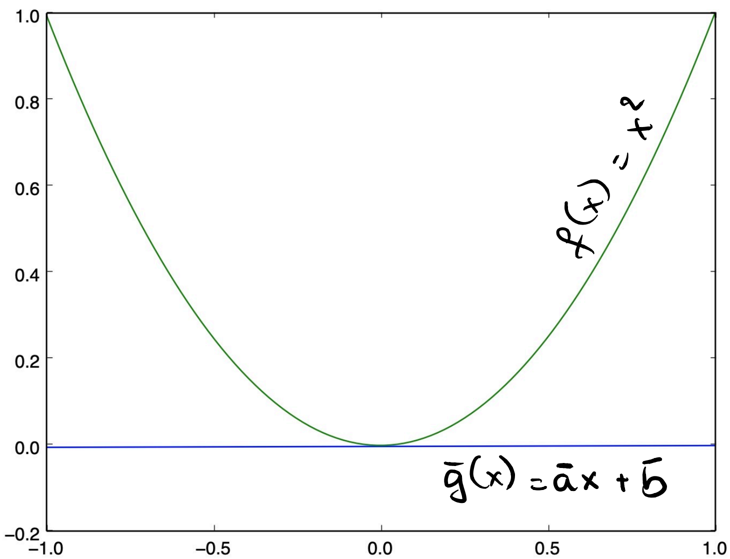
- E_{out} = mean ($E_{out}(g_i)$)

$$= \frac{1}{N} \sum_{i=1}^{N} \sum_{x=1}^{N} (a_i x + b_i - x^2)^2$$

- $\text{bias} = \frac{1}{N} \sum_{i=1}^{N} \left(\bar{g}(x_i) - f(x_i) \right)^2$

- $\text{var} = E_x \left[\frac{1}{N} \sum_{i=1}^{N} \left(g(x_i) - \bar{g}(x_i) \right)^2 \right.$

c,



$E_{out} = 0.53$

$\text{bias} = 0.2$

$\text{var} = 0.33$

$\rightarrow E_{out} = \text{bias} + \text{var}$

$\bar{g}(x) \approx 0$

d,

- $E_{out} = E_x\left[(g(x) - f(x))^2 \right] = E_x\left[(ax + b - x^2)^2 \right]$

$= E_x\left[x^4 - 2ax^3 + a^2x^2 - 2bx^2 + 2abx + b^2 \right]$

$= E_x[x^4] - 2aE_x[x^3] + (a^2 - 2b)E_x[x^2] + 2abE_x[x] + b^2$

$= \frac{1}{2}\int_{-1}^{1} x^4 dx - 2a\frac{1}{2}\int_{-1}^{1} x^3 dx + \frac{a^2 - 2b}{2}\int_{-1}^{1} x^2 dx + 2ab\int_{-1}^{1} x dx + b^2$

$$= \frac{1}{2}\left[\frac{1}{5}x^5\right]_{-1}^{1} - a\left[\frac{1}{4}x^4\right]_{-1}^{1} + \frac{a^2-2b}{2}\left[\frac{1}{3}x^3\right]_{-1}^{1} + 2ab\left[\frac{1}{2}x^2\right]_{-1}^{1} + b^2$$

$$= \frac{1}{5} - 0 + \frac{2}{3}\left(\frac{a^2-2b}{2}\right) + 0 + b^2$$

$$= \frac{a^2-2b}{3} + b^2 + \frac{1}{5}$$

$$\hookrightarrow E_D[E_{out}] = \frac{1}{5} + \frac{1}{3}E_D[(x_1+x_2)^2 + 2x_1x_2] + E_D[x_1^2x_2^2]$$

$$= \frac{1}{5} + \frac{1}{3}\times\frac{1}{4}\int_{-1}^{1}\int_{-1}^{1}(x_1^2+x_2^2+4x_1x_2)\,dx_1dx_2 + \frac{1}{4}\int_{-1}^{1}\int_{-1}^{1}x_1^2x_2^2\,dx_1dx_2$$

$$= \frac{1}{5} + \frac{1}{3}\times\frac{1}{4}\times\frac{8}{3} + \frac{1}{4}\times\frac{4}{9} = \boxed{\frac{8}{15}}$$

- bias $= E_x[(\bar{g}(x) - f(x))^2] = E_x[f(x)^2]$ $\quad(\bar{g}(x) \simeq 0)$

$$= E_x[x^4] = \frac{1}{2}\int_{-1}^{1}x^4\,dx = \boxed{\frac{1}{5}}$$

- var $= E_x[E_D[(g^{(D)}(x) - \underbrace{\bar{g}(x)}_{=0})^2]]$

$$= E_x[E_D[(ax+b)^2]] = E_x[E_D[a^2x^2 + 2axb + b^2]]$$

$$= E_x[E_D[(x_1+x_2)^2].x^2 - 2E_D[(x_1+x_2)x_1x_2]x + E_D[x_1^2x_2^2]]$$

$$= E_x\Big[\frac{1}{4}\int_{-1}^{1}\int_{-1}^{1}(x_1^2 + 2x_1x_2 + x_2^2)\,dx_1\,dx_2 \cdot x^2$$

$$- \frac{2}{4}\int_{-1}^{1}\int_{-1}^{1}\left(x_1^2x_2 + x_1x_2^2\right)dx_1\,dx_2 \cdot x$$

$$+ \frac{1}{4}\int_{-1}^{1}\int_{-1}^{1} x_1^2x_2^2\,dx_1\,dx_2$$

$$= E_x\left[\frac{2}{3}x^2 + \frac{1}{9}\right] = \frac{2}{3}\left(\frac{1}{2}\int_{-1}^{1}x^2dx\right) + \frac{1}{9} = \boxed{\frac{1}{3}}$$

4) LFD Problem 3.4

a)

$$E_n(w) = (\max(0, 1 - y_n w^T x_n))^2 = \begin{cases} 0 & , 1 - y_n w^T x_n \leqslant 0 \\ (1 - y_n w^T x_n)^2 & , 1 - y_n w^T x_n > 0 \end{cases}$$

When $E_n(w) = 0 \rightarrow \lim\limits_{1 - y_n w^T x_n \to 0^{\pm}} = 0 \rightarrow$ continuous

$$\nabla E_n(w) = \begin{cases} 0 & , 1 - y_n w^T x_n \leqslant 0 \\ -2 y_n x_n (1 - y_n w^T x_n) & , 1 - y_n w^T x_n > 0 \end{cases}$$

$\rightarrow \lim\limits_{1 - y_n w^T x_n \to 0^{\pm}} \nabla E_n(w) = 0 \rightarrow$ differentiable

$$\Rightarrow \quad \nabla E_n(w) = 0$$

b)

- $\text{sign}(w^T x_n) \neq y_n \rightarrow y_n w^T x_n \leqslant [0, 1)$

$\hookrightarrow [[\text{sign}(w^T x_n) \neq y_n]] = 1$

$y_n w^T x_n \leqslant 0 \rightarrow E_n(w) = (1 - y_n w_T x_n)^2 \geqslant 1$

$$\Rightarrow \quad E_n(w) \geqslant [[\text{sign}(w^T x_n) \neq y_n]]$$

- $\text{sign}(w^T x_n) = y_n \rightarrow y_n w^T x_n \geq 0$

$\hookrightarrow [[\text{sign}(w^T x_n) \neq y_n]] = 0 \rightarrow y_n w^T x_n \geq 0$

$$\hookrightarrow E_n(w) = \begin{cases} (1 - y_n w^T x_n)^2 \geq 0 & \text{if } 0 \leq y_n w^T x_n < 1 \\ 0 & \text{if } y_n w^T x_n \geq 1 \end{cases}$$

$$\Rightarrow E_n(w) \geq [[\text{sign}(w^T x_n) \neq y_n]]$$

$\therefore$ $E_n(w)$ is an upper bound for $[[\text{sign}(w^T x_n) \neq y_n]]$

$$\rightarrow \frac{1}{N} \sum_{n=1}^{N} E_n(w) \geq \frac{1}{N} \sum_{n=1}^{N} [[\text{sign}(w^T x_n) \neq y_n]] = E_{in}(w)$$

c) ADALINE : $w \leftarrow w + \eta(o - y)x$

Select (x_n, y_n) randomly

- If $y_n w^T x_n < 1 \rightarrow \nabla E_n(w) = -2y_n(1 - y_n w^T x_n)x_n$

$$\rightarrow w \leftarrow w - \eta \nabla E_n(w) = w + 2\eta y_n (1 - y_n w^T x_n) x_n$$
$$= w + 2\eta (y_n - \underbrace{y_n^2}_{=1} w^T x_n) x_n$$
$$= w + 2\eta (y_n - w^T x_n) x_n$$

- If $y_n w^T x_n > 1 \rightarrow \nabla E_n(w) = 0$

$$\Rightarrow w \leftarrow w - \eta \nabla E_n(w) = w - \eta . 0 = w$$

5, LFD Problem 3.19

a,

- This feature maps each point not in the dataset to the origin
- The complexity goes to infinity as $n \to \infty$

b,

- The complexity goes to infinity as $n \to \infty$

↳ $\phi_n(x)$ would be computationally expensive

c,

For $N < 10^4$, this will not generalize well