Cairo University - Faculty of Engineering
Computer Engineering Department
Machine Intelligence CMP 402B – Spring 2022

Assignment 2

Muhammad Alaa Abdel-Khaleq Sec: 2 BN: 22

April 2022

Problem 2.24

(a) $E_{in}(g) = \sum_{i=1}^{2} (f(x_i) - h(x_i))^2 = \sum_{i=1}^{2} (x_i^2 - (a x_i + b))^2$

① $\frac{\partial E_{in}(g)}{\partial a} = -2 \sum_{i=1}^{2} x_i (x_i^2 - a x_i - b) = 0$

② $\frac{\partial E_{in}(g)}{\partial b} = -2 \sum_{i=1}^{2} (x_i^2 - a x_i - b) = 0$

x_1 ① $-$ x_2 ②

$x_1^2 - a x_1 - b = 0$

$x_2^2 - a x_2 - b = 0$

$a = x_1 + x_2$

$b = -x_1 x_2$

so $g^D(x) = (x_1 + x_2) x - x_1 x_2$

$\bar{g}(x) = E_D(g^D(x)) = E_D((x_1 + x_2) x - x_1 x_2)$

$= E_D[x_1] x + E_D[x_2] x - E_D[x_1] E_D[x_2]$

duo to independence of x_1, x_2

since data follows a uniform [-1, 1]

$E_D[x_1] = 0$ $E_D[x_2] = 0$

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

b)

1. get $\bar{g}(x)$

- fix x
- for a number of times, e.g 1000
  - Sample two data points from $[-1, 1]$
  - Compute $g^D(x)$ using a, b derived In last question
- ~~Take the avr~~ Take the average value of $g^D(x)$ so we get $\bar{g}(x)$ at x

2. get Variance and E_{out}, bais

- for a number of times, e.g 5000
  - Sample x from $[-1, 1]$
  - follow the procedure to get $\bar{g}(x)$ to generate an arry of values of function $g^D(x)$ evaluated at that x
  - Compute the variance $E_D[(g^D(x) - \bar{g}(x))^2]$
  - we will use $\bar{g}(x)$ to compute $[(\bar{g}(x) - f(x))^2]$ at each x
  - we use the array of values to compute an array $(g^D(x) - \bar{g}(x))^2$. take the average of the resulting arry. we get $E_D[(g^D(x) - f(x))^2]$
- now we take the average of above calculated $E_D[(g^D(x) - \bar{g}(x))^2]$, $[(\bar{g}(x) - f(x))^2]$, $E_D[(g^D(x) - f(x))^2]$ and get the expected values of var, bais, E_{out}

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

Exercise 3.7

we take derivative of $E_{in}(w)$ with respect to w, $E_{in}(w) = \frac{1}{N}\sum_{1}^{N} \ln(1+e^{-y_n w^T x_n})$

$$\nabla E_{in}(w) = -\frac{1}{N}\sum_{1}^{N} \frac{y_n x_n e^{-y_n w^T x_n}}{1+e^{-y_n w^T x_n}}$$

$$= \frac{1}{N}\sum_{1}^{N} -y_n x_n \theta(-y_n w^T x_n)$$

when a sample is misclassified $y_n w^T x_n < 0$, $\theta(-y_n w^T x_n) > 0.5$ and when a sample is correctly classified $\theta(-y_n w^T x_n) < 0.5$, so the contribution of misclassified example is more to the gradient than a correctly classified one.