



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

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

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

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

$$x_1 (1) - x_2 (2)$$

$$x_1^2 - ax_1 - b = 0$$

$$x_2^2 - ax_2 - b = 0$$

$$a = x_1 + x_2$$

$$b = -x_1 x_2$$

$$\text{So } g^D(x) = (x_1 + x_2)x - x_1 x_2$$

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

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

due to independence of x_1, x_2

Since data follows 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 $g^p(x)$ using a, b derived in last question
- ~~Take the average~~ Take the average value of $g^p(x)$ so we get $\bar{g}(x)$ at x

2. get Variance and Entropy

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

Exercise 3.7

we take derivative of $E_{in}(w)$ with respect to w , $E_{in}(w) = \frac{1}{N} \sum_1^N \ln(1 + e^{-y_n w^T x_n})$

$$\begin{aligned}\nabla E_{in}(w) &= -\frac{1}{N} \sum_1^N \frac{y_n x_n e^{-y_n w^T x_n}}{1 + e^{-y_n w^T x_n}} \\ &= \frac{1}{N} \sum_1^N -y_n x_n \sigma(-y_n w^T x_n)\end{aligned}$$

when a sample is misclassified

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

3.12

$$H^2 = X(X^T X)^{-1} X^T X (X^T X)^{-1} X^T$$

$$H^2 = X(X^T X)^{-1} X^T \quad (1)$$

$$\begin{aligned} H^T &= (X(X^T X)^{-1} X^T)^T = X((X^T X)^T)^{-1} X^T \\ &= X(X^T X)^{-1} X^T = H \quad (2) \end{aligned}$$

from (1) & (2) H is a projection matrix

$\hat{y} = Hy \rightarrow \hat{y}$ is the projection of y onto the space spanned by X .