Cover page for answers.pdf
CSE512 Fall 2018 - Machine Learning - Homework 5

Your Name: Saif Suleman Vazir

Solar ID: 112072061

NetID email address: saifsuleman.vazir@stonybroook.edu

Names of people whom you discussed the homework with:

HOMEWORK 5

Q)1) i) Given:

$H(x) = \text{sign}(f(x))$

$f(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$

$\delta(H(x^j) \neq y^j) = 1$

$\delta(H(x^j) = y^j) = 0$

To prove:

$$\varepsilon_{Train} = \frac{1}{N} \sum_{j=1}^{N} \delta(H(x^j) \neq y^j) \leq \frac{1}{N} \sum_{j=1}^{N} e^{-f(x^j)y^j}$$

we know that for a classification task, training error is given by:

$$\varepsilon_{train} = \frac{1}{N} \sum_{j=1}^{N} \delta(H(x^j) \neq y^j)$$

Now, from the definition of $H(x)$ we get:

$$H(x) = \text{sign}(f(x)) = \begin{cases} 1 & f(x) > 0 \\ 0 & f(x) = 0 \\ -1 & f(x) < 0 \end{cases}$$

$\therefore$ cases when $H(x^j) \neq y^j$:

i) $y^j = +1$ $\quad H(x^j) = 0, -1$ $\quad \{ f(x^j) = 0$ or $f(x^j) < 0 \}$

ii) $y^j = -1$ $\quad H(x^j) = 0, 1$ $\quad \{ f(x^j) = 0$ or $f(x^j) > 0 \}$

$\therefore$ we can see that for the above cases:

$y^j \cdot H(x^j) \leq 0$ or $y^j \cdot f(x^j) \leq 0$.

$\therefore$ Our new training error $\Rightarrow$

$$\varepsilon_{train} = \frac{1}{N} \sum_{j=1}^{N} E(x) \quad \text{where } E(x) = \begin{cases} 1 & \text{when } y^j f(x^j) \leq 0 \\ 0 & y^j f(x^j) > 0 \end{cases}$$

$\therefore$ ~~If we map this function to:~~

~~$e^{-E(x)}$, we get:~~

~~$e^{-E(x)} \geq 1$ when $E(x) \leq 0$~~

Now considering $e^{-y^j f(x^j)}$ we get :

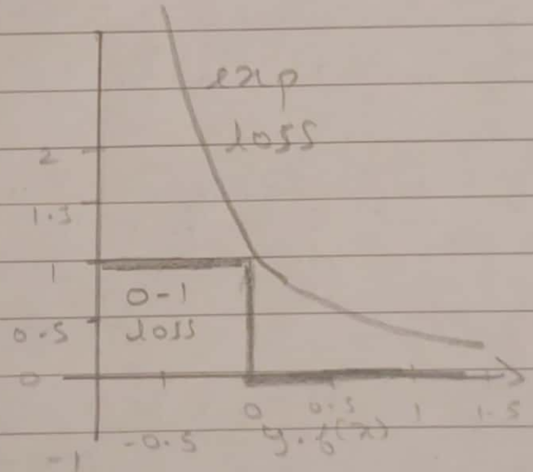
$$e^{-y^j f(x^j)} = \begin{cases} \geq 1 & y^j f(x^j) \leq 0 \\ 0 < z \leq 1 & y^j f(x^j) > 0 \end{cases}$$

[Proof of training error in question 1.2]

Hence, from the above eq^{n}s we get that

$$\frac{1}{N} \sum_j \delta(H(x^j) \neq y^j) \leq \frac{1}{N} \sum_j e^{-y^j f(x^j)}$$

This can also be visualized graphically as left hand side is 0-1 loss and right hand side is exponential loss f^n. And we know that 0-1 loss is always less than or equal to exponential loss.



Q1) 2) We have:

$$w_j^{(t+1)} = \frac{w_j^{(t)} e^{-\alpha_t y^j h_t(x^j)}}{Z_t}$$

$$\therefore w_j^{t} = \frac{w_j^{(t-1)} e^{-\alpha_{t-1} y^j h_{t-1}(x^j)}}{Z_{t-1}}$$

$$\vdots$$

$$w_j^{k} = \frac{w_j^{(k-1)} e^{-\alpha_{k-1} y^j h_{k-1}(x^j)}}{Z_{k-1}}$$

Also we know that:
$w_j^{(1)} = 1/N$

$\therefore$ From the above eq^s we get:

$$w_j^{(t+1)} = w_j^{(1)} \cdot \frac{e^{-\alpha_1 y^j h_1(x^j)}}{Z_1} \cdot \frac{e^{-\alpha_2 y^j h_2(x^j)}}{Z_2} \cdots \frac{e^{-\alpha_t y^j h_t(x^j)}}{Z_t}$$

$$= \frac{1}{N} \cdot \frac{1}{\prod_{i=1}^{t} Z_i} \cdot e^{-y^j(\alpha_1 h_1(x^j) + \alpha_2 h_2(x^j) \cdots \alpha_t h_t(x^j))}$$

$$w_j^{(t+1)} = \frac{1}{N} \frac{e^{-y^j \sum_{t=1}^{t} \alpha_t h_t(x^j)}}{\prod Z_t}$$

As $\sum \alpha_t h_t(x^j) = f(x^j)$

$$\therefore w_j^{(t+1)} = \frac{1}{N} \frac{e^{-y^j f(x^j)}}{\prod Z_t}$$

$$\therefore \frac{1}{N} e^{-y^j f(x^j)} = w_j^{(t+1)} \prod_{t=1}^{T} Z_t$$

$\therefore$ Summing for all values of $j = 1$ to N

$$\frac{1}{N} \sum_{j=1}^{N} e^{-y^j f(x^j)} = \sum_{j=1}^{N} w_j^{(t+1)} \cdot \prod Z_t$$

$$\frac{1}{N} \sum_{j=1}^{N} e^{-y^j f(x^j)} = \prod_{t=1}^{T} Z_t \quad \left[\text{as } \sum_{j=1}^{N} w_j^{(t+1)} = 1 \right]$$

Q1) 3) Given $Z_t = (1-\varepsilon_t) e^{-\alpha_t} + \varepsilon_t e^{\alpha_t}$

Now, in order to minimize Z_t:

a) $$\frac{\partial Z_t}{\partial \alpha_t} = \frac{\partial \left[(1-\varepsilon_t) e^{-\alpha_t} + \varepsilon_t e^{\alpha_t} \right]}{\partial \alpha_t}$$

$$= (1-\varepsilon_t) e^{-\alpha_t}(-1) + \varepsilon_t e^{\alpha_t} = 0$$

$$\varepsilon_t e^{\alpha_t} = (1-\varepsilon_t) e^{-\alpha_t}$$

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

$$2\alpha_t = \ln \frac{1-\varepsilon_t}{\varepsilon_t} \qquad \therefore \alpha_t = \frac{1}{2} \ln \frac{1-\varepsilon_t}{\varepsilon_t}$$

Scanned by CamScanner

$$\therefore Z_t^{\alpha_t} = (1-\varepsilon_t)\, e^{-\frac{1}{2}\ln\frac{1-\varepsilon_t}{\varepsilon_t}} + \varepsilon_t\, e^{\frac{1}{2}\ln\frac{1-\varepsilon_t}{\varepsilon_t}}$$

$$= (1-\varepsilon_t)\cdot\sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}} + \varepsilon_t\cdot\sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}$$

$$= \sqrt{(1-\varepsilon_t)\varepsilon_t} + \sqrt{\varepsilon_t(1-\varepsilon_t)}$$

$$\therefore Z_t^{\alpha_t} = 2\sqrt{(1-\varepsilon_t)\varepsilon_t}$$

b) substituting $\varepsilon_t = 1/2 - \gamma_t$ in the above eq^n we get

$$Z_t = 2\sqrt{(1-1/2+\gamma_t)(1/2-\gamma_t)}$$

$$= 2\sqrt{(1/2+\gamma_t)(1/2-\gamma_t)}$$

$$= 2\sqrt{\frac{1}{4}-\gamma_t^2}$$

$$= 2\sqrt{\frac{1-4\gamma_t^2}{4}} = \sqrt{1-4\gamma_t^2}$$

Taking log we get:

$$\log Z_t = \frac{1}{2}\{\log 1-4\gamma_t^2)$$

$$\cancel{= \frac{1}{2}[\log 1-2\gamma_t + \log 1+2\gamma_t]}$$

$$\cancel{\le \frac{1}{2}[-2\gamma_t - 2\gamma_t]}$$ [using $\log 1-x \le -x$ & $\log 1+x \le x$] $x = 2\gamma_t$, $x = -2\gamma_t$

$$\therefore \cancel{\log Z_t \le -2\gamma_t}$$

$$\log Z_t \le \frac{1}{2}[-4\gamma_t^2]$$ [using $\log 1-x \le -x$] $4\gamma_t^2$

$$\therefore \log Z_t \le -2\gamma_t^2$$

$$\therefore Z_t \le e^{-2\gamma_t^2}$$

c) $\epsilon_{train} \leq \prod_{t=1}^{T} z_t \leq \prod_{t=1}^{T} e^{-2\gamma_t^2}$

$\leq e^{-2\sum_{t=1}^{T}\gamma_t^2}$

As we have that:

$\gamma_t \geq \gamma \qquad (\gamma > 0)$

squaring both sides:

$\gamma_t^2 \geq \gamma^2$

summing over all t:

$\sum_{t=1}^{T} \gamma_t^2 \geq \sum_{t=1}^{T} \gamma^2$

$\sum_{t=1}^{T} \gamma_t^2 \geq T\gamma^2$

Multiplying by -2 we get:

$-2\sum_{t=1}^{T} \gamma_t^2 \leq -2T\gamma^2$

Exponentiation:

$e^{-2\sum_t \gamma_t^2} \leq e^{-2T\gamma^2}$

From the eqⁿ for training error we get:

$\epsilon_{train} \leq e^{-2\sum_t \gamma_t^2} \leq e^{-2T\gamma^2}$

Hence proved.

2.5.1

Total sum within groups, [p1,p2,p3]:

K=2

5.3648 * 1.0e+08, [79.82   54.81   67.31]

K=4

4.6111* 1.0e+08, [67.88   86.83   77.36]

K=6

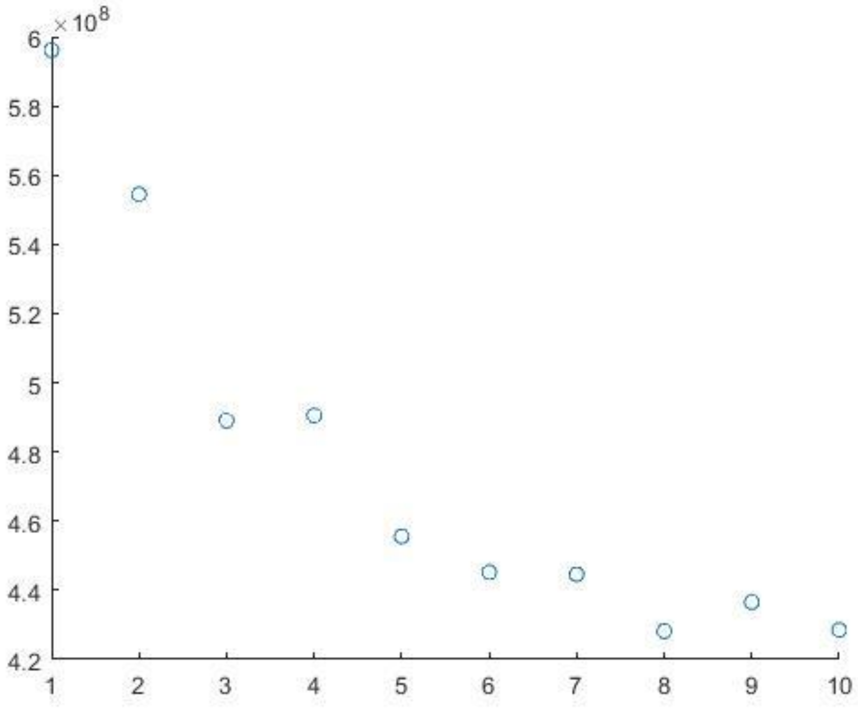
4.3135* 1.0e+08, [55.18   94.43   74.81]

2.5.2

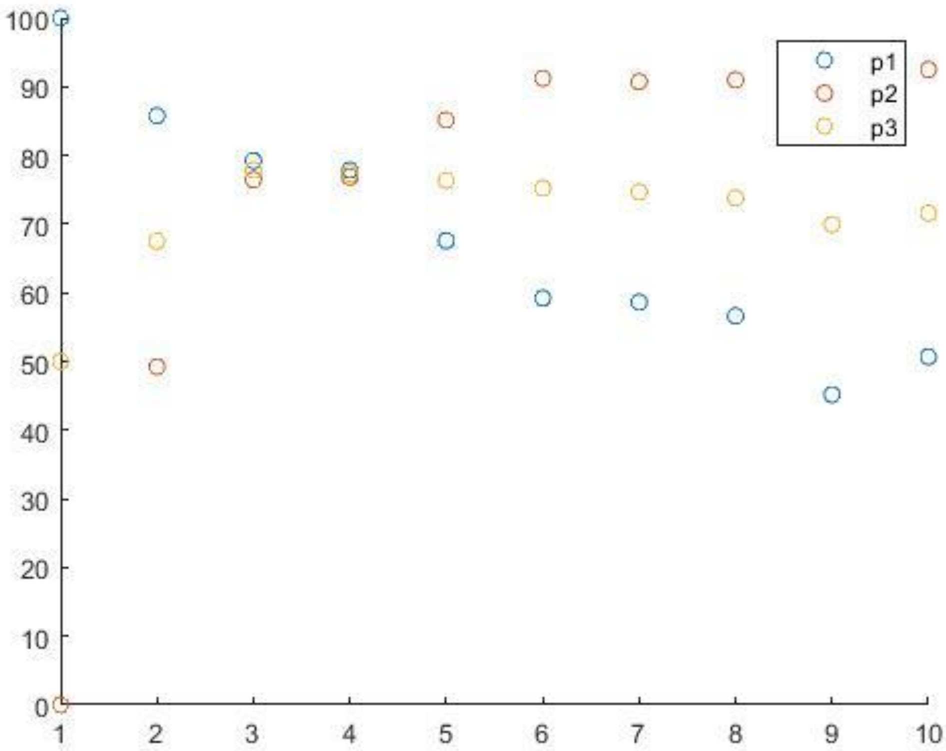
Number of iterations = 8

2.5.3

Sum of squares vs k



2.5.4



3.4.2

Accuracy with 5-fold CV = 15.6443

3.4.3

Accuracy with 5-fold CV = 84.018

C=100

Gamma=20

3.4.5

Accuracy with 5-fold CV=90.2082

C=200

Gamma=20