CSE512 Fall 2019 - Machine Learning - Homework 5

November 19, 2019

Name: Anmol Shukla

Solar ID: 112551470

NetID: anmshukla

Email address: Anmol.shukla@stonybrook.edu

Names of people whom I have discussed the homework with:

- Abhiraj Smit - Asmit@cs.stonybrook.edu

Question 1

1. To Prove:

$$\epsilon_{Training} = \frac{1}{N}\sum_{j=1}^{N}\delta\left(H(x^j) \neq y^j\right) \leq \frac{1}{N}\sum_{j=1}^{N}\exp(-f(x^j)y^j)$$

The condition for error is: $H(x^j) \neq y^j$

where,

$$H(x^j) = sgn\left\{\sum_{t=1}^{T}\alpha_t h_t(x)\right\} = sgn\{f(x)\}$$

i.e. $H(x^j)$ is the predicted label for i^{th} data point
And the actual label is $y^j \in \{-1, 1\}$

The two cases when an error can occur is:-

(1) When $H(x^j) = 1$ & $y^j = -1$

$\therefore$ $sgn\{f(x^j)\} = 1$

$\therefore$ $f(x^j) \geq 0$

$\therefore$ $y^j f(x^j) < 0$

(2) When $H(x^j) = -1$ & $y^j = 1$

$sgn\{f(x^j)\} = -1$

$\therefore$ $f(x^j) < 0$

$\therefore y^j f(x^j) < 0$

$\therefore$ $\boxed{y^j f(x^j) < 0}$

Similarly, in case of no error, $y^j f(x^j) > 0$

$$\epsilon_{Training} = \frac{1}{N}\sum_{j=1}^{N}\begin{cases}1 & , \text{if } H(x^j) \neq y^j \\ 0 & , \text{if } H(x^j) = y^j\end{cases}$$

$$= \frac{1}{N}\sum_{j=1}^{N}\begin{cases}1 & \text{if } f(x^j)y^j < 0 \\ 0 & \text{if } f(x^j)y^j > 0\end{cases}$$

In the worst case scenario, all the N training data points could contribute to the error, i.e. $f(x^j).y^j < 0$ for $\forall j$.

$$\therefore \quad \epsilon_{Training} = \frac{1}{N} \sum_{j=1}^{N} (1) = \frac{1}{N} \times N = 1 \quad \text{(in worst case)}$$

$$\& \quad \epsilon_{Training} = \frac{1}{N} \sum_{j=1}^{N} (0) = 0 \quad \text{(in best case)}$$

$$\therefore \quad 0 \leq \epsilon_{Training} \leq 1$$

We know that

$$f(x^j) y^j \leq 0$$

$$-f(x^j) y^j \geq 0$$

taking exponent on both the sides

$$e^{-f(x^j)y^j} \geq 1$$

$$\sum_{j=1}^{N} e^{-f(x^j).y^j} \geq \sum_{j=1}^{N} (1)$$

$$\sum_{j=1}^{N} e^{-f(x^j).y^j} \geq N$$

$$\frac{1}{N} \sum_{j=1}^{N} e^{-f(x^j).y^j} \geq 1$$

& we already proved that $\epsilon_{Training} \leq 1$

$$\therefore \quad \epsilon_{Training} \leq \frac{1}{N} \sum_{j=1}^{N} e^{-f(x^j).y^j}$$

2 To Prove:

$$\frac{1}{N}\sum_{j=1}^{N} \exp\left(-f(x^j)y^j\right) = \prod_{t=1}^{T} Z_t$$

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

$t=1$

$$w_j^{(2)} = \frac{w_j^{1} \exp\left(-\alpha_1 y^j h_1(x^j)\right)}{Z_1}$$

$t=2$

$$w_j^{(3)} = \frac{w_j^{(2)} \exp\left(-\alpha_2 y^j h_2(x^j)\right)}{Z_2}$$

$$w_j^{(3)} = \frac{w_j^{1} e^{-\alpha_1 y^j h_1(x^j)}}{Z_1} \times \frac{\cancel{w_j^{2}}\, e^{-\alpha_2 y^j h_2(x^j)}}{Z_2}$$ (Substituting value of $w_j^{(2)}$)

In general, we can write $w_j^{(t+1)}$ as:

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

$$w_j^{(t+1)} = \frac{w_j^{(t)} \, e\left(-y^j \sum_{t=1}^{T} \alpha_t h_t(x^j)\right)}{\prod_{t=1}^{T} Z_t} \quad -①$$

Weights at any step t should sum upto 1 i.e. $\sum_{j=1}^{N} w_j^{(t)} = 1$

$\therefore$ for first time-step, each $w_j^{(1)} = \frac{1}{N}$ as $\sum_{j=1}^{N} w_j^{(1)} = 1$

eq^n ① can be written as $\quad$ since $\sum_{t=1}^{T} \alpha_t h_t(x^j)$

$$w_j^{(t+1)} = \frac{w_j^{(t)} \, e^{-y^j f(x^j)}}{\prod_{t=1} Z^t}$$

But we know that $\sum_{j=1}^{N} w_j^{(t+1)} = 1$

$\therefore$ our previous equation can be written as:

$$\sum_{j=1}^{N} w_j^{(t+1)} = \frac{1}{N} \frac{\sum_{j=1}^{N} e^{-y^j f(x^j)}}{\prod_{t=1}^{T} Z_t}$$ (applying Σ on both sides)

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

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

Hence Proved

3.
(a)

$$Z_t = (1-\epsilon_t)e^{-\alpha_t} + \epsilon_t e^{\alpha_t}$$

to minimize Z_t, we'll differentiate w.r.t. α

$$\frac{dZ_t}{d\alpha_t} = -(1-\epsilon_t)e^{-\alpha_t} + \epsilon_t e^{\alpha_t} = 0$$

$$\cancel{\epsilon} - \cancel{e^{-\alpha_t}} + \cancel{\epsilon_t e^{-\alpha_t}} + \cancel{\epsilon_t e^{\alpha_t}} = 0$$

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

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

$$e^{\alpha_t} = \sqrt{\frac{1-\epsilon_t}{\epsilon_t}}$$

$$\alpha_t = \ln\left(\sqrt{\frac{1-\epsilon_t}{\epsilon_t}}\right)$$

for Z_t^{opt} we'll substitute this value of α_t in Z_t

$$Z_t^{opt} = (1-\epsilon_t)\exp\left(-\ln\sqrt{\frac{1-\epsilon_t}{\epsilon_t}}\right) + \epsilon_t \exp\left(\ln\sqrt{\frac{1-\epsilon_t}{\epsilon_t}}\right)$$

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

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

$$Z_t^{opt} = 2\sqrt{\epsilon_t(1-\epsilon_t)}$$

Hence Proved

(b) Given :- $\epsilon_t = \frac{1}{2} - \gamma_t$

We already proved that $Z_t = Z_t^{opt} = 2\sqrt{\epsilon_t(1-\epsilon_t)}$

$$\therefore Z_t = 2\sqrt{\left(\tfrac{1}{2}-\gamma_t\right)\left(1-\tfrac{1}{2}+\gamma_t\right)}$$

$$= 2\sqrt{\left(\tfrac{1}{2}\right)^2 - \gamma_t^2}$$

$$= \sqrt{1-4\gamma^2} \quad -① $$

We have also been given that -

$$\log(1-x) \leq -x \quad \text{for } 0 \leq x \leq 1$$

$$(1-x) \leq e^{-x}$$

$$\sqrt{1-x} \leq e^{-x/2} \quad -②$$

$$\epsilon_t \geq 0$$

$$\therefore \tfrac{1}{2} - \gamma_t \geq 0$$

$$\gamma_t \leq \tfrac{1}{2}$$

$$4\gamma_t^2 \leq 1 \quad -③$$

$$\gamma_t^2 \geq 0$$

$$4\gamma_t^2 \geq 0 \quad -④$$

from eqn ①, ② ③, ④

$$Z_t \leq e^{-2\gamma_t^2}$$

$$\therefore \prod_{t=1}^{T} Z_t \leq \prod_{t=1}^{T} e^{(-2\gamma_t^2)}$$

$$\prod_{t=1}^{T} Z_t \leq e^{(-2\sum_{t=1}^{T}\gamma_t^2)}$$

$$\therefore \epsilon_{Training} \leq \prod_{t=1}^{T} Z_t \leq e^{-(2\sum_{t=1}^{T}\gamma_t^2)}$$

Hence Proved

(c) To prove:

$$\epsilon_{Training} \leq e^{-2T.\gamma^2}$$

we know that,

$$\gamma_t \geq \gamma \quad \forall t, \gamma > 0$$

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

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

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

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

taking exponent on both sides

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

we already know that, $\epsilon_{Training} \leq e^{-2\sum_{t=1}^{T}\gamma_t^2}$

$\therefore$ combining the above 2 eqn $\rightarrow$ $\epsilon_{Training} \leq e^{-2T\gamma^2}$

$\therefore$ As T increases, $\epsilon_{Training}$ decreases

2 Programming Question (clustering with K-means)

2.5.1

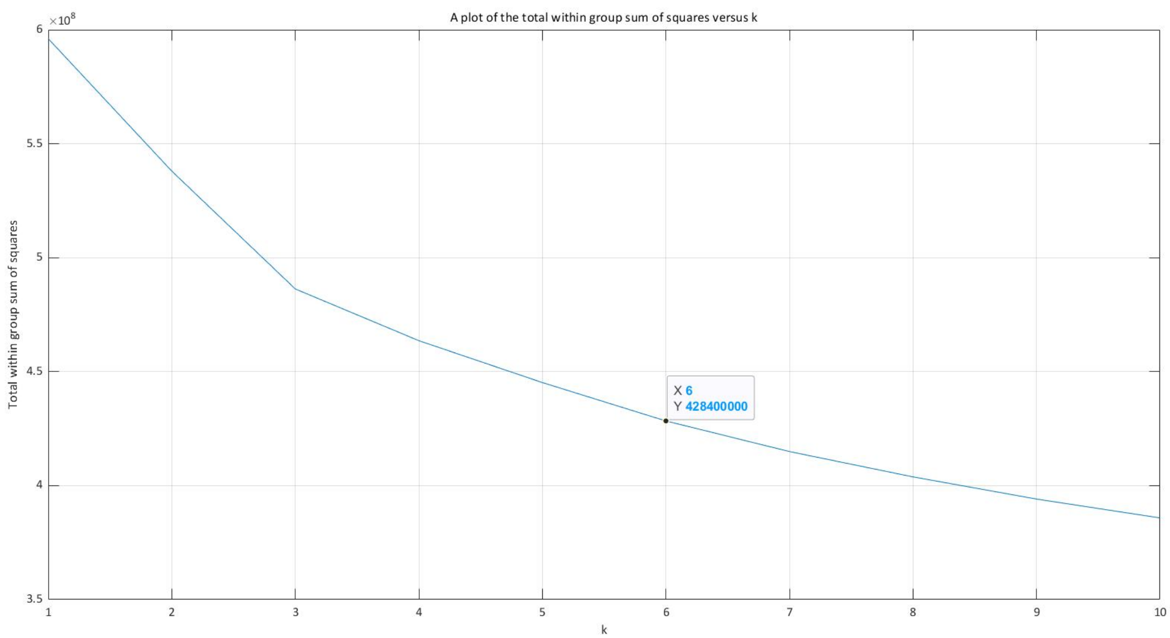
k	Sum of Squares	p1	p2	p3
2	5.3676 x 1.0e+08	79.815689	54.805460	67.310575
4	4.6111 x 1.0e+08	67.881215	86.832918	77.357066
6	4.3135 x 1.0e+08	55.176546	94.434978	74.805762

2.5.2

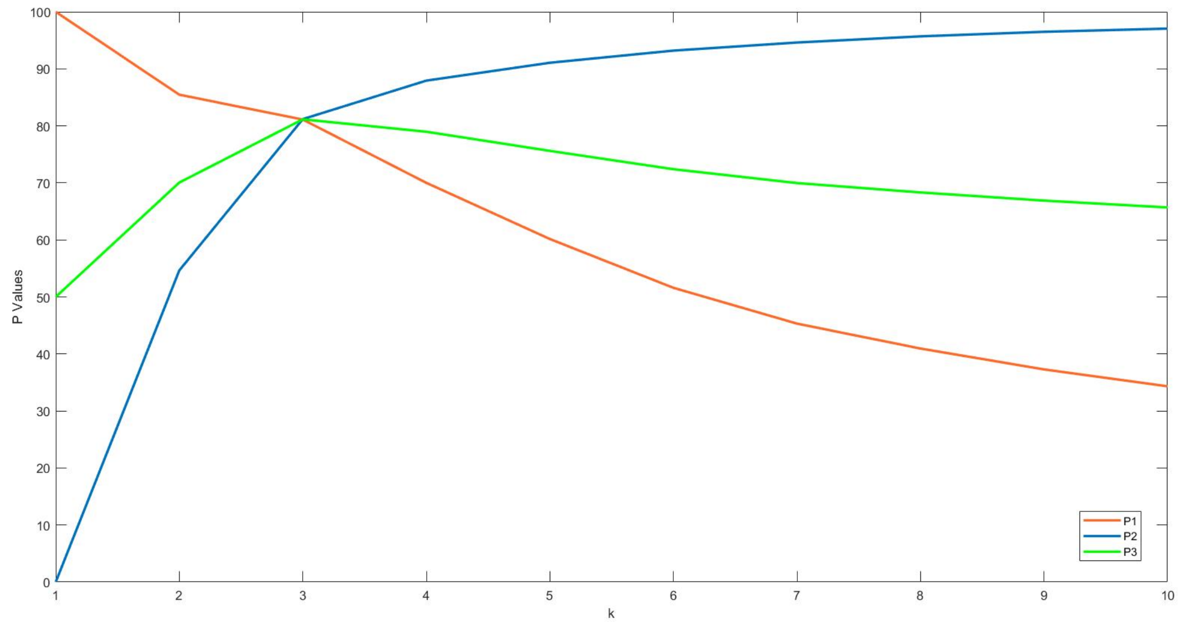
The number of iterations that k-means ran for k = 6 is 8.

2.5.3

Sum of squares versus k for k = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10



2.5.4



3 Programming Question (scene classification)

3.4.1

Please see the corresponding matlab file for HW5 BoW.learnDictionary().

3.4.2

The accuracy for 5-fold CV with default parameters = 15.64

3.4.3

After tuning:
$C = 8192$
$\gamma = 8$
Accuracy = 89.45

3.4.4

Please see the corresponding code in exponentialKernel.m file.

3.4.5

After tuning:
$C = 2048$
$\gamma = 2$
5-fold CV Accuracy = 94.01

3.4.6

Kaggle score = 0.80208