CSE512 Fall 2018 - Machine Learning - Homework 5

Name: Rohit Bhal
Solar ID: 112073893
NetID email address: rbhal@cs.stonybrook.edu / rohit.bhal@.stonybrook.edu

1. We are given :

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

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

$$\delta(H(x^j) \neq y^j) = 1 \quad , \quad H(x^j) \neq y^j$$
$$0 \quad , \quad \text{otherwise}$$

Now, we have four conditions for y^j & $f(x^j)$:

(1) $y^j = 1$ & we predict $f(x^j) = 1$
$y^j f(x^j) = 1$

(2) when we have true, $y^j = -1$ & we predict $f(x^j) = 1$.
$y^j f(x^j) = -1$

(3) when we have true, $y^j = 1$ & we predict $f(x^j) = -1$.
$y^j f(x^j) = -1$.

(4) when we have true, $y^j = -1$ & we predict $f(x^j) = -1$
$y^j f(x^j) = 1$.

we can see from the above that $y^j f(x^j) \geq 0$ means no error & $y^j f(x^j) \leq 0$ means error exists.

Now, $\epsilon_{train} = \frac{1}{N} \sum_{j=1}^{N} \begin{cases} 1 & , \text{ If } y^j f(x^j) \leq 0 \\ 0 & , \text{ otherwise} \end{cases}$

For some z, $\exp(-z) \geq 1$ if $z \leq 0$
$\implies \exp(-y^j f(x^j)) \geq 1$ if $y^j f(x^j) \leq 0$

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

2 We have:

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

where Z_t is a normalizing constant ensuring that the weights sum to 1.

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

We will consider for 1 sample & find out the value of $w_j^{(t+1)}$:

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

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

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

$$= w_j^{(t-1)} \frac{\exp(-\alpha_t y^j h_t(x^j))}{Z_t} \frac{\exp(-\alpha_{t-1} y^j h_{t-1}(x^j))}{Z_{t-1}}$$

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

But $w_j = w_i = 1/N_i = \frac{1}{N_i}$ (only 1 sample)

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

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

$$\therefore w_j^{(t+1)} = \frac{1}{N_i} \exp(-y^j f(x^j))$$

Now, we can extend the same for N data points:

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

Now, since Z_t is a normalizing constant.

$$\therefore \sum_{j=1}^{N} w_j^{(t+1)} = 1$$

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

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

3 $\quad \epsilon_t = \sum_{j=1}^{N} w_j^{(t)} \delta(h_t(x^j) \neq y^j)$

$$Z_t = (1-\epsilon_t)\exp(-\alpha_t) + \epsilon_t \exp(\alpha_t)$$

(a) $\quad Z_t^{opt} = 2\sqrt{\epsilon_t(1-\epsilon_t)}$ To Prove

$$\frac{\partial(Z_t)}{\partial \alpha} = 0 = (1-\epsilon_t)\exp(-\alpha_t)(-1) + \epsilon_t \exp(\alpha_t)$$

$$\exp(\alpha_t)\,\epsilon_t = (1-\epsilon_t)\exp(-\alpha_t)$$

$$\frac{(1-\epsilon_t)}{\epsilon_t} = (\exp(\alpha_t))^2$$

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

$$\alpha_t = \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)^{1/2}$$

Put α_t value,

$$Z_t = (1-\epsilon_t)\left(\frac{1-\epsilon_t}{\epsilon_t}\right)^{-1/2} + \epsilon_t\left(\frac{1-\epsilon_t}{\epsilon_t}\right)^{1/2}$$

$$= ((1-\epsilon_t)\epsilon_t)^{1/2} + (\epsilon_t(1-\epsilon_t))^{1/2}$$

$$= 2\sqrt{(1-\epsilon_t)\epsilon_t}.$$

Proved.

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

But, $\epsilon_t = \frac{1}{2} - \gamma_t$

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

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

$$= 2\sqrt{\left(\tfrac{1}{4}-\gamma_t^2\right)}$$

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

But $\log(1-x) \leq -x$

$\therefore$ $Z_t \leq \exp(-2\gamma_t^2)$

we need $\log(1-x) \leq -x$

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

$$\sqrt{1-x} \leq e^{-x/2}$$

Now $x = 4\gamma_t^2$

$$\therefore Z_t \leq e^{-\frac{4\gamma_t^2}{2}}$$

$$Z_t \leq \exp(-2\gamma_t^2) \quad \text{Proved}$$

(c) We know that $\gamma_t > 0$ implies that it is better than random.

$\therefore$ $\gamma_t \geq \gamma$, where $\gamma > 0$ $\forall t$

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

$$\Longrightarrow \epsilon_{train} \leq \exp\left(-2\sum_{t=1}^{T}\gamma^2\right)$$

$$\epsilon_{train} \leq \exp(-2T\gamma^2) \quad \text{Proved.}$$

2.5

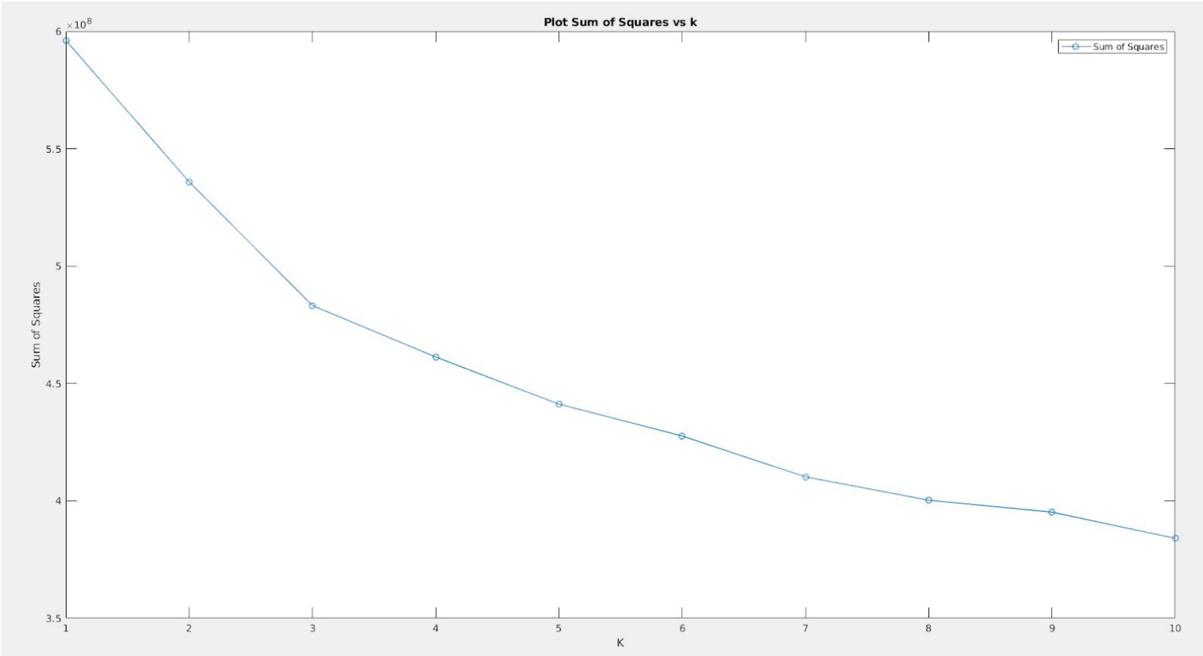
1.
k=2
Sum of Square(SS) = 5.3648 X 10^8
p1 =0.7982 p2=0.5481 p3=0.6731

k=4
SS=4.6111 X 10^8
p1=0.6788 p2=0.8683 p3=0.7736

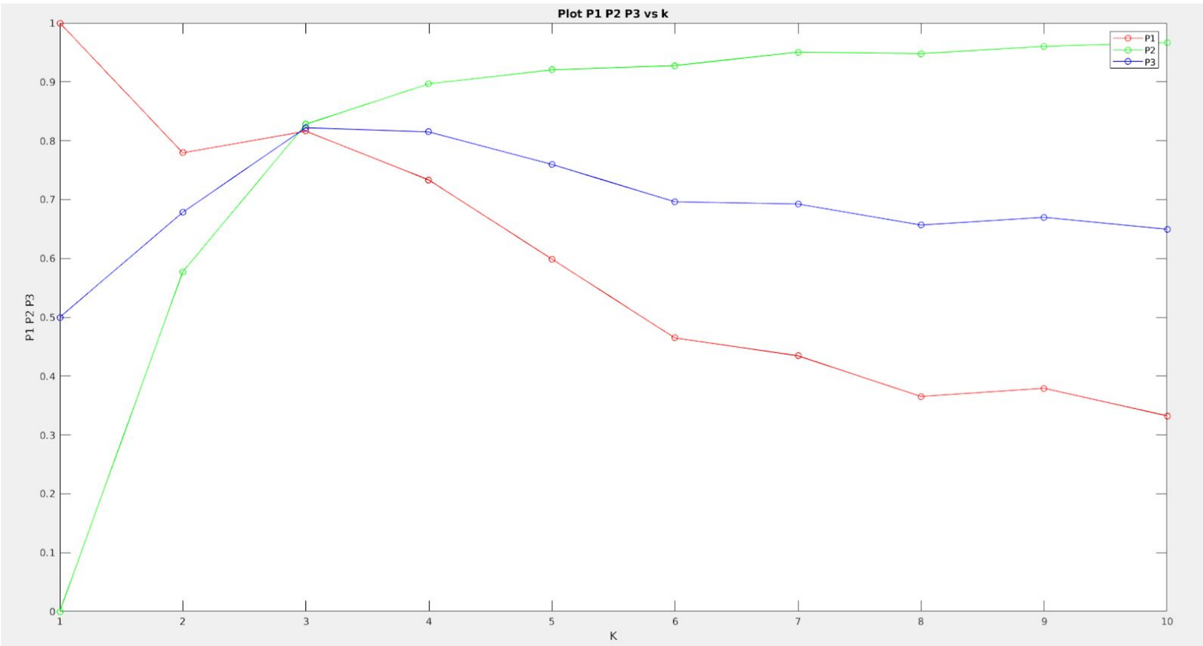
k=6
SS=4.3135 X 10^8
p1=0.5518 p2=0.9443 p3=0.7481

2.
The number of iterations that k-means ran for k=6 is 8 rounds.

3.


Plot Sum of Squares vs k
×10^8
Sum of Squares
Sum of Squares
K
6
5.5
5
4.5
4
3.5
1
2
3
4
5
6
7
8
9
10

4


Plot P1 P2 P3 vs k
P1
P2
P3
P1 P2 P3
K
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10

3.4

2.
Gamma = 0.001;
C=1;
Using RBS Kernel:
acc = svmtrain(trLbs, trDK, sprintf('-t 4 -v 5 -q'));
Accuracy = 15.6443%

3.
RBS GridSearch Using Libsvm svmtrain function
model = svmtrain(trLbs, trDK, sprintf('-t 4 -c %f -g %f -v 5 -q'));

	c=0.1	1	10	40	80	160
g=0.1	15.6443 %	15.6443 %	24.8734 %	44.6258 %	53.067 %	64.4344 %
1	15.6443 %	24.7046 %	58.188 %	70.6809 %	75.2392 %	79.1221 %
10	22.5661 %	58.4131 %	76.8711 %	84.4682 %	87.0568 %	87.5633 %
20	32.9207 %	65.8976 %	81.5419 %	86.9443 %	87.6759 %	88.1823 %
30	35.3405 %	68.8239 %	83.7366 %	87.6759 %	88.52%	88.7451 %
40	41.3618 %	70.6246 %	84.5808 %	87.6759 %	88.7451 %	88.5763 %
80	51.8852 %	75.5768 %	87.4508 %	88.1823 %	88.2949 %	88.2949 %
100	53.8548 %	76.5898 %	87.4508 %	87.901 %	87.9572 %	87.9572 %

5.

For Chi Square Kernel :
g = 1.400000   c = 20.000000
Cross Validation Accuracy = 93.5847%

6.
g = 10, c = 80
Accuracy from Kaggle = 83.250%