Cover page for answers.pdf CSE512 Fall 2019 - Machine Learning - Homework 3

Name	Vamshi Muthineni
Solar ID	112607946
NetID email address	vmuthineni@cs.stonybrook.edu
Names of people whom you discussed the homework with	AravindReddy Ravula, Rajireddy Annadi, Mallikarjuna, Vamshi krishna, Meghana, Fathima Cherat

1.1.1

(1.1.1) As we know that the optimal Bayes risk for data point x is

$$r^*(x) = \min_i \left\{ \sum_{i=1}^{M} \hat{n}_i(x) L(i,j) \right\}$$

$L(i,j)$ is the loss value when a data point is predicted as 'j' instead of actual class 'i'.

It is given for us that

		0	1
			True class i
Assigned j	0	0	α
	1	1	0

0 - positive
1 - Negative

Loss matrix.

$\therefore L(0,1) = 1$ & $L(1,0) = \alpha$.

when $j=0$

$$r^*(x) = \sum_{i=0}^{1} \hat{n}_i(x) \cdot L(i,j)$$

$$= \hat{n}_0(x) \cdot L(0,0) + \hat{n}_1(x) \cdot L(1,0)$$

$$= \hat{n}_0(x) \cdot (0) + (1-\eta(x)) \cdot \alpha$$

$$= (1-\eta(x))\alpha.$$

$\eta(x)$ is prob. that x is positive
So $(1-\eta(x))$ is prob. that x is -ve

when $j=1$

$$= \sum_{i=0}^{1} \hat{n}_i(x) \cdot L(i,1)$$

$$= \hat{n}_0(x) \cdot L(0,1) + \hat{n}_1(x) \cdot L(1,1)$$

$$= \eta(x) \cdot (1) + 0 = \eta(x).$$

So $\boxed{r^*(x) = \min\{ \eta(x) + (1-\eta(x)) \cdot \alpha \}.}$

1.1.2

(1.1.2) Asymptotic risk

$r(x)$ can be given by

$$r(x) = P(Y=1)\cdot P(Y_{neighbor}=0) + \alpha \cdot P(Y=0)\cdot P(Y_{neighbor}=1)$$

$$= \eta(x)\cdot\left[1-\eta(x)\right] + \alpha\cdot(1-\eta(x))\cdot\eta(x)$$

hence

$$\cancel{r(x) = \eta(x) - \eta(x)\cdot\eta(x) + \alpha\,\eta(x)}$$

$$r(x) = \eta(x)\cdot\left[1-\eta(x)+\alpha-\alpha\cdot\eta(x)\right]$$

$$= \eta(x)\cdot\left[(1-\eta(x))\cdot(1+\alpha)\right]$$

$$\boxed{r(x) = \eta(x)\cdot(1-\eta(x))\cdot(1+\alpha)}$$

1.1.3

(1.1.3) We know that

$$\gamma^*(x) = \min\{\eta(x), (1-\eta(x))\cdot\alpha\}.$$

case:1

consider $\eta(x)$ is maximum in above two terms.

i.e. $\eta(x) \geq \alpha(1-\eta(x))$

$$\eta(x) + \alpha\eta(x) \geq \alpha.$$

$$\eta(x) \geq \frac{\alpha}{1+\alpha}.$$

we know that $\eta(x) \leq 1$

hence

$$\frac{\alpha}{1+\alpha} \leq \eta(x) \leq 1. \quad \text{— (1)}$$

$$\gamma^*(x) = \alpha(1-\eta(x)).$$

$$\eta(x) = 1 - \frac{\gamma^*(x)}{\alpha} \quad \text{— (2)}$$

substituting (2) in (1)

$$\frac{\alpha}{1+\alpha} \leq 1 - \frac{\gamma^*(x)}{\alpha} \leq 1$$

$$\frac{\gamma^*(x)}{\alpha} \leq \frac{1}{1+\alpha}.$$

$$\frac{\gamma^*(x)}{\alpha}\cdot(1+\alpha) \leq 1$$

$$\gamma^*(x)\left(1+\frac{1}{\alpha}\right) - 1 \leq 0 \quad \text{— (3)}.$$

from 1.1.2 we've

$$\gamma(x) = (1+\alpha)\eta(x)(1-\eta(x))$$

$$= (1+\alpha)\left(1-\frac{\gamma^*(x)}{\alpha}\right)\cdot\left(\cancel{1}-\cancel{1}+\frac{\gamma^*(x)}{\alpha}\right) \quad \text{from (2).}$$

$$= (1+\alpha)\left(1-\frac{\alpha\gamma^*(x)}{\alpha}\right)\cdot\left(\frac{\gamma^*(x)}{\alpha}\right) \quad \text{— (4)}$$

Now consider

$$\left(1-\frac{\gamma^*(x)}{\alpha}\right)\cdot\left(\frac{\gamma^*(x)}{\alpha}\right)-(1-\gamma^*(x))\cdot\gamma^*(x)$$

$$=\left(\frac{1}{\alpha}-\frac{\gamma^*(x)}{\alpha^2}-1+\gamma^*(x)\right)\cdot\gamma^*(x)$$

$$=\left(\left(\frac{1}{\alpha}-1\right)+\gamma^*(x)\left(1-\frac{1}{\alpha^2}\right)\right)\gamma^*(x)$$

$$=\left(\gamma^*(x)\left(1+\frac{1}{\alpha}\right)-1\right)\cdot\gamma^*(x)\cdot\left(1-\frac{1}{\alpha}\right)$$

here $A \cdot B$.

$\gamma^*(x)\left(1+\frac{1}{\alpha}\right)-1$ is ≤ 0 from 3.

and $\gamma^*(x)\cdot\left(1-\frac{1}{\alpha}\right)>0$ since $1-\frac{1}{\alpha}>0$.

hence The above multiplication is ≤ 0.

$$\left(1-\frac{\gamma^*(x)}{\alpha}\right)\left(\frac{\gamma^*(x)}{\alpha}\right)-(1-\gamma^*(x))(\gamma^*(x))\leq 0.$$

$$\left(1-\frac{\gamma^*(x)}{\alpha}\right)\left(\frac{\gamma^*(x)}{\alpha}\right)\leq(1-\gamma^*(x))\gamma^*(x). \quad \text{— 5}$$

from 4 & 5

we get

$$\gamma(x)\leq(1+\alpha)(1-\gamma^*(x))\gamma^*(x) \quad \text{— 5}.$$

Case 2:

Now consider

$$\eta(x)<\alpha(1-\eta(x))$$

$$\eta(x)+\alpha\eta(x)<\alpha.$$

$$\eta(x)<\frac{\alpha}{1+\alpha}.$$

So $\gamma^*(x)=\eta(x)$ i.e. $\min\{\eta(x),\alpha(1-\eta(x))\}$.

Substituting in 1.1.2.

we get

$$\gamma(x) = (1+\alpha)\eta(x)(1-\eta(x))$$

$$= (1+\alpha)\gamma^*(x)(1-\gamma^*(x)) \quad \text{--- } 6 .$$

from 5 & 6 it is clelar that.

$$\boxed{\gamma(x) \leq (1+\alpha)\gamma^*(x)(1-\gamma^*(x))}$$

1.1.4

(1.1.4) Let R be the asymptotic risk of the 1-NN classifier & R^* be the bayes risk.

since

$$R(x) = E(r(x))$$

from 1.1.3

$$r(x) \le (1+\alpha) r^*(x)(1-r^*(x))$$

$$E(r(x)) \le E\left[(1+\alpha) r^*(x)(1-r^*(x))\right]$$

$$\le (1+\alpha) E\left[r^*(x) - r^*(x)^2\right]$$

$$\le (1+\alpha)\left[E(r^*(x)) - E(r^*(x)^2)\right]$$

$$\le (1+\alpha)\left[E(r^*(x)) - \left[var(r^*(x)) + [E(r^*(x))]^2\right]\right]$$

since $E(x^2) = var(x) + [E(x)]^2$

$$\le (1+\alpha)\left[E(r^*(x))(1-E(r^*(x))) - var(r^*(x))\right]$$

$$\le (1+\alpha)\left[R^*(1-R^*) - var(r^*(x))\right]$$

$$\le (1+\alpha) R^*(1-R^*) - (1+\alpha) var(r^*(x))$$

hence. $\boxed{R \le (1+\alpha) R^*(1-R^*)}$ since $1+\alpha \ge 0$, $var(r^*(x)) \ge 0$

1.2.1

(1.2)

(1.2.1) As we know that

Asymptotic risk in K-NN classifier

$$r(x) = P(Y=1) \cdot P(Y_{majority\text{-}neighbors} = 0) + P(Y=0) \cdot P(Y_{majority\text{-}neighbors} = 1)$$

given ~~from~~ in the data that

$$P(Y=1) = \eta(x), \quad P(Y_{majority\text{-}neighbors} = 1) = g(n,k)$$

hence $P(Y=0) = 1-\eta(x)$, $P(Y_{majority\text{-}neighbors} = 0) = 1-g(n,k)$

Substituting in the above eqn.

$$\boxed{r(x) = \eta(x) \cdot [1-g(n,k)] + (1-\eta(x)) \cdot g(n,k)}$$

1.2.2

(1.2.2) As we know

$$\gamma^*(x) = \min\{\eta(x), 1-\eta(x)\}$$

from the above derivation we've

$$\gamma(x) = \eta(x)\cdot(1-g(\eta,k)) + (1-\eta(x))\cdot g(\eta,k) \quad \text{—} \ (1)$$

So when $\eta(x) \le \frac{1}{2}$

$$\gamma^*(x) = \min\{\eta(x), 1-\eta(x)\}$$

$$= \eta(x)$$

i.e. $\eta(x) = \gamma^*(x)$

Substituting in (1)

$$\gamma(x) = \gamma^*(x)(1-g(\eta,k)) + (1-\gamma^*(x))g(\eta,k)$$

$$= \cancel{\gamma^*(x) - \gamma^*(x)\cdot g(\eta,k) +}$$

$$= \gamma^*(x) - \gamma^*(x)\cdot g(\eta,k) + (1-\gamma^*(x))\cdot g(\gamma^*(x),k)$$

$$\gamma(x) = \gamma^*(x) + (1-2\gamma^*(x))\cdot g(\gamma^*(x),k) \quad \text{—} \ (2)$$

when $\eta(x) \ge \frac{1}{2}$

$$\gamma^*(x) = 1-\eta(x) \Rightarrow \eta(x) = 1-\gamma^*(x)$$

Substituting in (1)

$$\gamma(x) = (1-\gamma^*(x))\cdot(1-g(\eta,k)) + (\cancel{1}-\cancel{1}+\gamma^*(x))\cdot g(\eta,k)$$

$$= 1-\gamma^*(x) + [\gamma^*(x) - 1 + \gamma^*(x)]\cdot g(\eta,k)$$

$$= 1-\gamma^*(x) + [2\gamma^*(x) - 1]\cdot g(1-\gamma^*(x),k)$$

$$= 1-\gamma^*(x) + (2\gamma^*(x)-1)\cdot[1-g(\gamma^*(x),k)]$$

$$\gamma(x) = \gamma^*(x) + (1-2\gamma^*(x))\cdot g(\gamma^*(x),k) \quad \text{—} \ (3)$$

1.2.3

So from ②, ③, it is clear for us that in any case

$$\boxed{\gamma(x) = \gamma^*(x) + (1-2\gamma^*(x))\cdot g(\gamma^*(x), K)}$$

(1.2.3) Hoeffding's inequality ~~pro~~ is

$$P(H(n)) \geq (p+\epsilon)n \leq \exp(-2\epsilon^2 n).$$

$H(n) \rightarrow$ no. of points that are positive

$p \rightarrow$ probability that a point is positive

n = total points.

$$g(\gamma^*(x), K) = \begin{cases} P(H(n) \geq \frac{K}{2}) & \text{when } n \text{ is even} \\ P(H(n) > \frac{K}{2} & \text{when } n \text{ is odd.} \end{cases}$$

So $P(H \geq \frac{K}{2}) = P(H \geq (\gamma^*(x) + 0.5 - \gamma^*(x))K)$.

observing with Hoeffdings inequality it's clear that $p = \gamma^*(x)$

$\epsilon = 0.5 - \gamma^*(x)$

$n = K.$

So we'll get

$$g(\gamma^*(x), K) = \boxed{P(H \geq \frac{K}{2}) \leq \exp(-2(0.5-\gamma^*(x))^2 K)}$$

1.2.4

(1.2.4)

We need to prove

$$\gamma(x) \le \gamma^*(x) + \frac{1}{\sqrt{2K}}.$$

from 1.2.2, we know that

$$\gamma(x) = \gamma^*(x) + (1-2\gamma^*(x))\, g(\gamma^*(x), K)$$

rewriting the equation as follows

$$\gamma(x) = \gamma^*(x) + \frac{1}{\sqrt{2K}} \cdot \left[\sqrt{2K}(1-2\gamma^*(x))\, g(\gamma^*(x), K)\right] \quad — (7)$$

from 1.2.3

$$g(\gamma^*(x), K) \le \exp\left(-2(0.5-\gamma^*(x))^2 K\right) \quad — (8)$$

substitute (8) in (7).

$$\gamma(x) \le \gamma^*(x) + \frac{1}{\sqrt{2K}} \left[\sqrt{2K}(1-2\gamma^*(x)) . \exp\left(-2(0.5-\gamma^*(x))^2 K\right)\right]$$

$$\le \gamma^*(x) + \frac{1}{\sqrt{2K}} \left[2\sqrt{2K}(0.5-\gamma^*(x)) . \exp\left(-2(0.5-\gamma^*(x))^2 K\right)\right]$$

Let $Z = \sqrt{2K}(0.5-\gamma^*(x))$

the the above eqn. will be of form

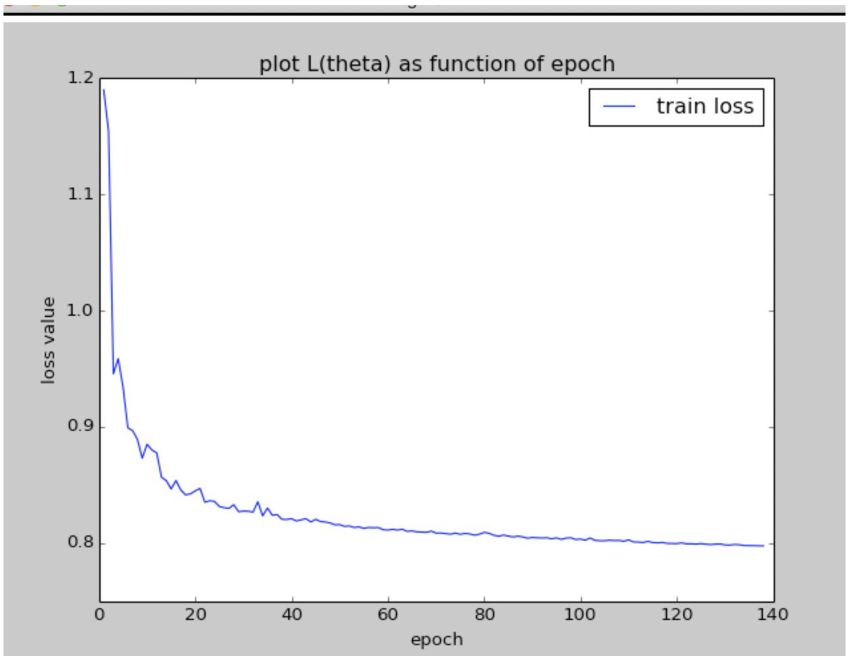
$$\le \gamma^*(x) + \frac{1}{\sqrt{2K}}\left[2Z\exp(-Z^2)\right]$$

The values for $2Z\exp(-Z^2)$ lies between $[-0.858, 0.858]$ which is ≤ 1.

hence $\boxed{\gamma(x) \le \gamma^*(x) + \frac{1}{\sqrt{2K}}}$

2.3.1

a. Number of epochs before exiting = 138
b. Plot showing L(theta) as a function of epoch

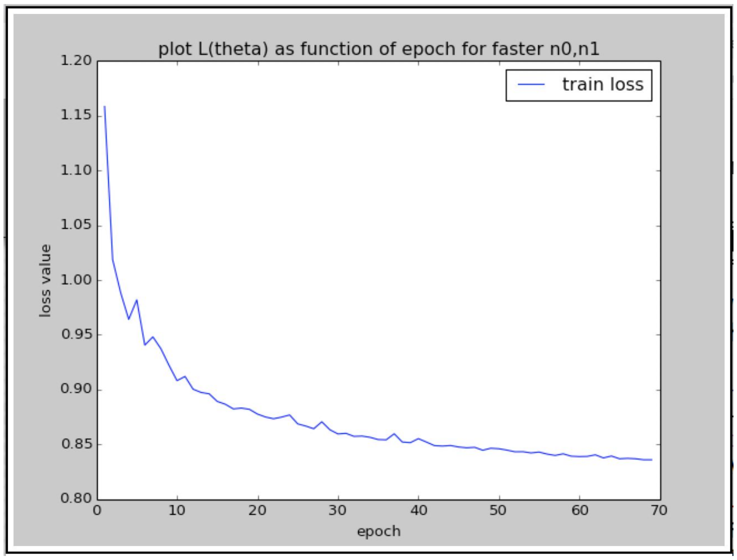


c. Final value of L(theta) after the optimization = 0.8

2.3.2.

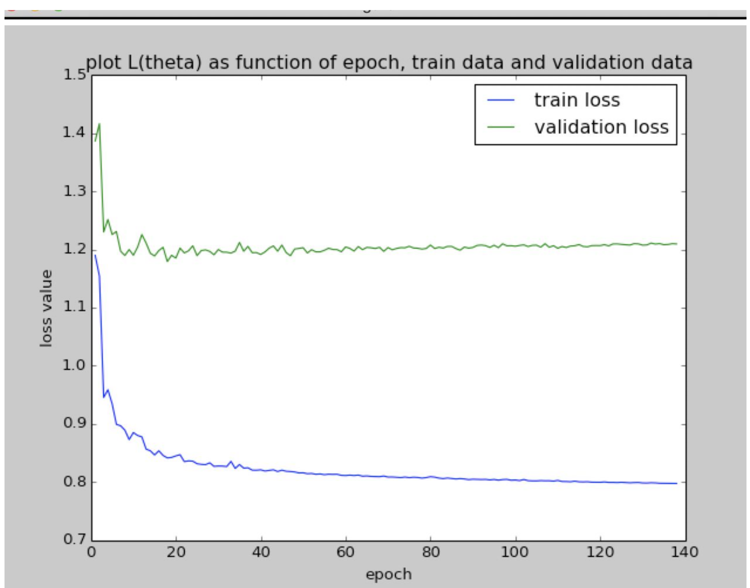
a. The pair of n0,n1 that leads to faster convergence is 0.1, 5.
   The no of epochs taken is 70.
   Final value of L(theta) = 0.85

b.

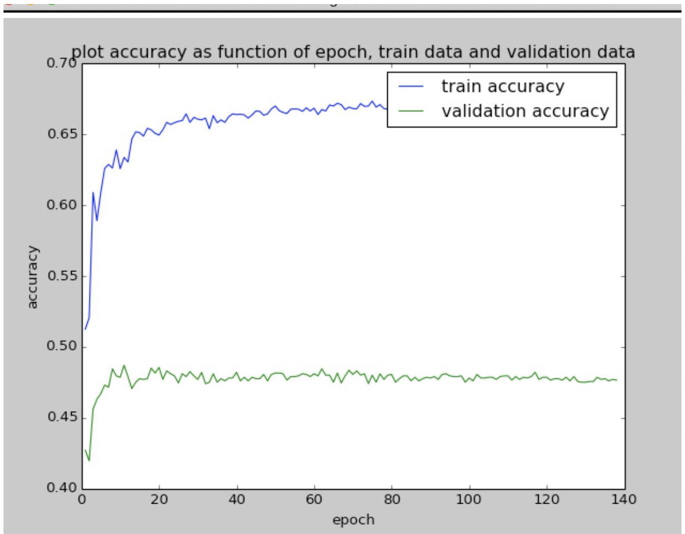

plot L(theta) as function of epoch for faster n0,n1
train loss
loss value
epoch
1.20
1.15
1.10
1.05
1.00
0.95
0.90
0.85
0.80
0
10
20
30
40
50
60
70

2.3.3

a.


plot L(theta) as function of epoch, train data and validation data
train loss
validation loss
loss value
epoch
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0
20
40
60
80
100
120
140

b.



2.3.4. Confusion matrix

On Train Data

		Predicted Values			
		1	2	3	4
	1	536	185	28	34
Actual	2	139	911	114	105
Values	3	40	231	353	192
	4	29	113	93	897

On Validation Data

		Predicted Values			
		1	2	3	4
	1	169	94	26	27
Actual	2	133	359	77	91
Values	3	39	197	77	125
	4	25	116	90	355

2.4

1. Best accuracy obtained on test data on kaggle - 46%