CSE512 Fall 2018 - Machine Learning - Homework 5

Your Name: Astitv Nagpal

Solar ID: 112008011

NetID email address: anagpal@cs.stonybrook.edu
Names of people whom you discussed the homework with: Ayush Garg

Question 1 – Boosting :

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

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

We have $(x_1, y_1), (x_2, y_2), (x_3, y_3) \ldots, (x_N, y_N)$ & $y_i \in \{-1, 1\}$

$h_t(x) \rightarrow$ weak classifier.

$h_t : X \rightarrow \{-1, 1\}$] $\rightarrow$ given

Therefore we have a distribution D_t.

→ Goodness of h_t ~~for~~ the distribution D_t is.

$$Pr_{D_t}(h_t(x_i) \neq y_i) = \sum_{h_t(x_i) \neq y_i} D_t$$

Next step D_t can be written as:

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

Also, we are given that y_i & classifier belongs to $\{-1, 1\}$. , we get :

$$D_{T+1}(i) = D_1(i) \left[\frac{\exp(-\alpha_1 y_i h_1(x_i))}{Z_1} \cdots \frac{\exp(-\alpha_T y_i h_T(x_i))}{Z_T}\right]$$

$$= \frac{1}{N} \frac{\exp\left(-y_i \sum_{t=1}^{T} \alpha_t h_t(x_i)\right)}{\prod_{t=1}^{T} Z_t} = \frac{1}{N} \frac{\exp(-y_i f(x_i))}{\prod_{t=1}^{T} Z_t} \longrightarrow \text{(ii)}$$

using eq^n (i)

Also given that,

$H(x_i) \neq y_i$ then $y_i f(x_i) \leq 0$

This gives us :

$$\exp(-y_i f(x_i)) \geq 1$$ ∴ using this we get.

$$H(x_i) \neq y_i \leq \exp(-y_i f(x_i))$$

⟶ summing over all is we get

$$\frac{1}{N}\sum_{i=1}^{T}(H(x_i) \neq y_i) \leq \frac{1}{N}\sum_{i=1}^{T}\exp(-y_i f(x_i)) \quad \text{— ⓐ}$$

HENCE PROVED

② ~~Substituting eqn (ii) in (a) we get.~~

② Using eqn ⓐ we get.

$$\frac{1}{N}\sum_{i=1}^{T}(H(x_i) \neq y_i) \leq \frac{1}{N}\sum_{i=1}^{T}\exp(-y_i f(x_i))$$

By substituting eqn (ii) here we get

$$\frac{1}{N}\sum_{i=1}^{T}(H(x_i) \neq y_i) \leq \frac{1}{N}\sum_{i=1}^{T}\exp(-y_i f(x_i))$$

$$D_{T+1}(i) = \frac{1}{N}\exp(-y_i f(x_i)) \Big/ \prod_{t=1}^{T} Z_t$$

∴ using these equations we get

$$\frac{1}{N}\sum_{i=1}^{T}(H(x_i) \neq y_i) = \sum_{i=1}^{T} D_{T+1}(i) \prod_{t=1}^{T} Z_t$$

Here, since ~~dist~~ D_{T+1} is a distribution & we know that it sums to 1 ie. $\sum_{i=1}^{T} D_{T+1} = 1$ —— using this we get.

$$\frac{1}{N}\sum_{i=1}^{T}(H(x_i) \neq y_i) = \prod_{t=1}^{T} Z_t$$

Hence Proved

3

a) Given $Z_t = (1-\varepsilon_t)\exp(-\alpha_t) + \varepsilon_t \exp(\alpha_t)$ we perform a partial To find α_t that minimizes Z_t derivative w.r.t. α_t.

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

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

$$\varepsilon_t e^{2\alpha_t} = (1-\varepsilon_t) \Rightarrow e^{2\alpha_t} = \left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$$

∴ Taking ln on both side.

$$2\alpha_t = \ln\left[\frac{1-\varepsilon_t}{\varepsilon_t}\right] \Rightarrow \alpha_t = \frac{1}{2}\ln\left[\frac{1-\varepsilon_t}{\varepsilon_t}\right]$$

Scanned by CamScanner

To find Z_t^{opt} we substitute the value of α

We get.

$$= (1-\varepsilon_t) e^{-\frac{1}{2}\ln\left[\frac{1-\varepsilon_t}{\varepsilon_t}\right]} + \varepsilon_t e^{\frac{1}{2}\ln\left[\frac{1-\varepsilon_t}{\varepsilon_t}\right]}$$

$$= (1-\varepsilon_t)\left[\frac{1-\varepsilon_t}{\varepsilon_t}\right]^{-1/2} + (\varepsilon_t)\left[\frac{1-\varepsilon_t}{\varepsilon_t}\right]^{1/2}$$

$$= (1-\varepsilon_t)\left[\frac{\varepsilon_t}{1-\varepsilon_t}\right]^{1/2} + \varepsilon_t\left[\frac{1-\varepsilon_t}{\varepsilon_t}\right]^{1/2}$$

$$= \left[\frac{(1-\varepsilon_t)^2\,\varepsilon_t}{\cancel{1-\varepsilon_t}}\right]^{1/2} + \left[\frac{\varepsilon_t^2(1-\varepsilon_t)}{\cancel{\varepsilon_t}}\right]^{1/2}$$

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

$$= 2((1-\varepsilon_t)(\varepsilon_t))^{1/2} = \boxed{2\sqrt{\varepsilon_t(1-\varepsilon_t)} = Z_t^{opt}}$$

Hence Proved.

(b) $\varepsilon_t = \frac{1}{2} - \gamma_t$ —— (a)

we have $Z_t^{opt} = 2\left[\varepsilon_t(1-\varepsilon_t)\right]^{1/2}$ —— (i)

Substituting (a) in eq^n (i) we get.

$$= 2\left[\left(\tfrac{1}{2} - \gamma_t\right)\left(1 - \left(\tfrac{1}{2} - \gamma_t\right)\right)\right]^{1/2}$$

$$= 2\left[\left(\tfrac{1}{2} - \gamma_t\right)\left(\tfrac{1}{2} + \gamma_t\right)\right]^{1/2}$$ using $(a+b)(a-b) = a^2 - b^2$

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

$$= \left[1 - 4\gamma_t^2\right]^{1/2}$$ —— (ii)

We are give that $\log(1-x) \leq -x$ using exp on both sides

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

$\Rightarrow (1-x) \leq e^{-x}$ putting sq. root on both side

$(1-x)^{1/2} \leq e^{-\frac{1}{2}x}$ $\longrightarrow$ comparing with eq^n (ii) we get

$Z_t \leq e^{-\frac{1}{2}4\gamma_t^2} \Rightarrow \boxed{Z_t \leq e^{-2\gamma_t^2}}$

Hence Proved.

(c) From part (b) we get:

$$\mathcal{E}_{training} \le \prod_{t=1}^{T} Z_t \le \exp\left(-2\sum_{t=1}^{T} \gamma_t^2\right) \quad \text{(i)}$$

If each classifier is better than random i.e.

$\gamma_t \ge \gamma$, then we have.

$$= \exp -2\left(\sum_{t=1}^{T} \gamma_t^2\right) \quad \text{where } \gamma_t \Leftrightarrow \gamma.$$

$$= \exp\left(-2 \sum_{t=1}^{T} \gamma^2\right) \quad \text{taking } \gamma \text{ out as it's a constant.}$$

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

$\therefore$ using eqn (i) we get

$$\boxed{\mathcal{E}_{training} \le \exp\left(-2T\gamma^2\right)}$$

Hence Proved!

Question 2.5

1)

```
For K = 2
P1 = 8.402299e+01
P2 = 5.732043e+01
P3 = 7.067171e+03
Total SS Error = 5.395215e+08


For K = 4
P1 = 7.611253e+01
P2 = 7.190292e+01
P3 = 7.400772e+03
Total SS Error = 4.610922e+08


For K = 6
P1 = 6.900537e+01
P2 = 7.880544e+01
P3 = 7.390540e+03
Total SS Error = 4.348008e+08
```

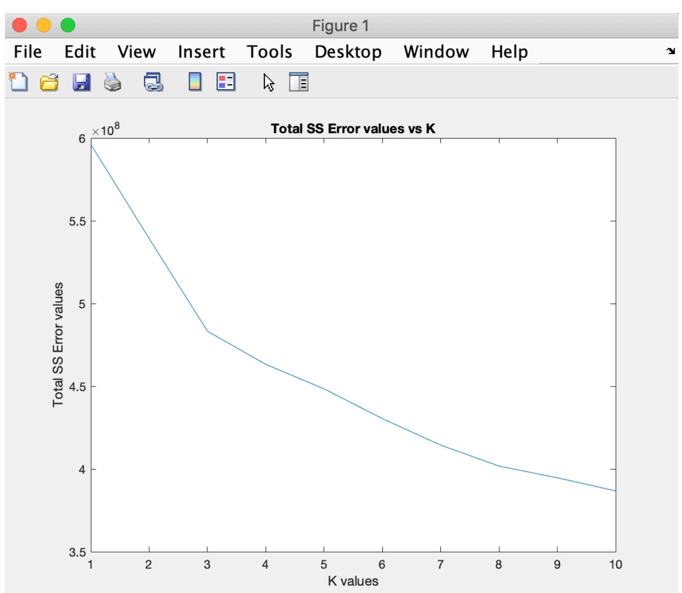
2)

```
K = 2
Total Iterations = 11

K = 4
Total Iterations = 17

K = 6
Total Iterations = 20
```

3)



4)

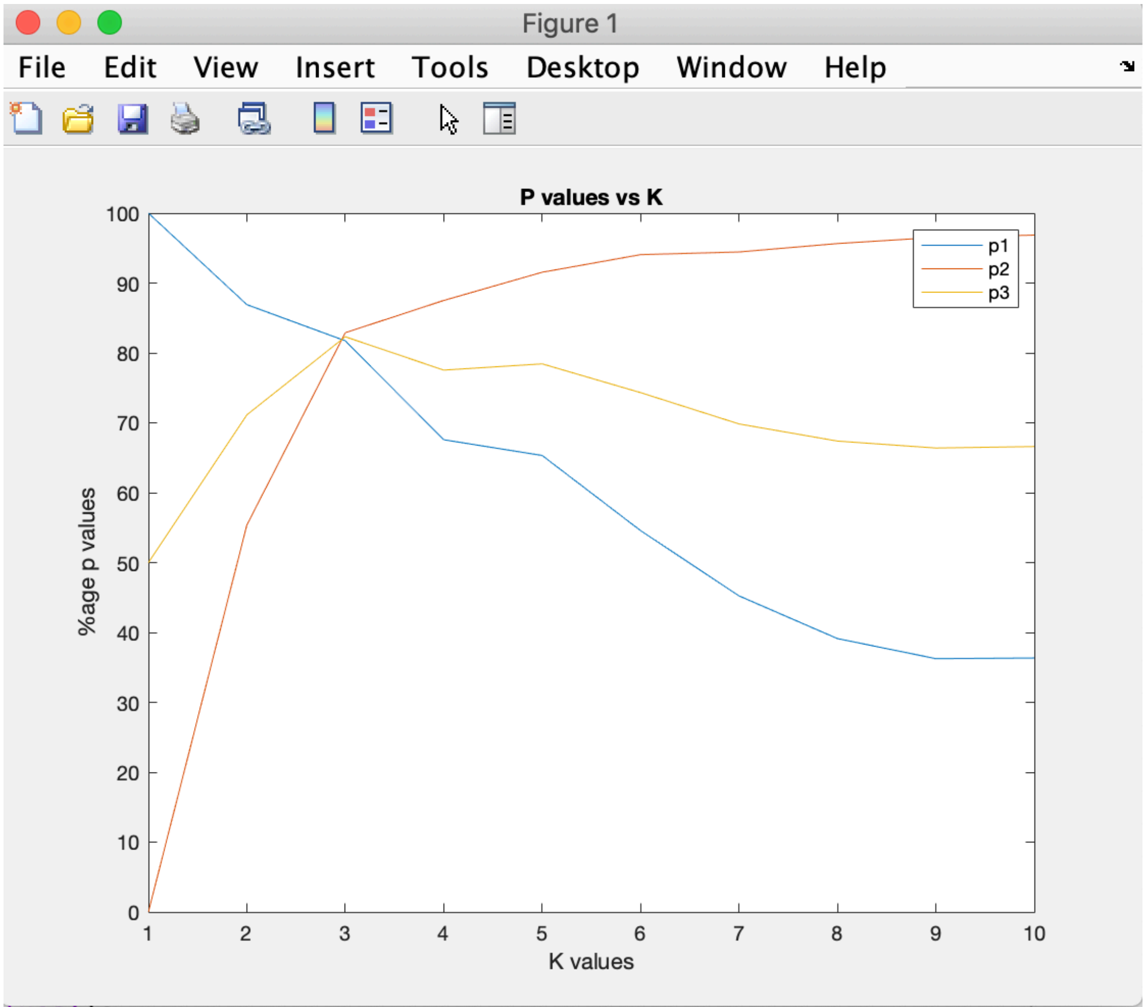


Figure 1
File Edit View Insert Tools Desktop Window Help
P values vs K
p1
p2
p3
%age p values
K values

Question 3.4

2)

```
optimization finished, #iter = 169
nu = 0.864450
obj = -337.995645, rho = -0.999944
nSV = 338, nBSV = 338
Total nSV = 1418
Cross Validation Accuracy = 15.6443%
   15.6443
```

3)

C = 10000
Gamma = 1

```
optimization finished, #iter = 2362
nu = 0.040672
obj = -79519.717269, rho = 11.748406
nSV = 103, nBSV = 0
Total nSV = 824
Cross Validation Accuracy = 89.0827%
   89.0827
```

4)

```
function [trainKFinal, testKFinal] = cmpExpX2Kernel(trainD, testD)
[trDSize,~] = size(trainD);
[tstDSize, ~] = size(testD);
trainK = zeros(trDSize, trDSize);
testK = zeros(tstDSize, trDSize);

for i = 1:trDSize
    num = trainD - trainD(i,:);
    den = trainD + trainD(i,:);
    trainK(:,i) = sum(num.^2 ./ (den+eps), 2);
end

trainGamma = mean(trainK, 'all');
trainK = exp(trainK*(-1/trainGamma));

for i = 1:trDSize
    num = testD - trainD(i,:);
    den = testD + trainD(i,:);
    testK(:,i) = sum(num.^2 ./ (den+eps), 2);
end

testGamma = mean(testK, 'all');
testK = exp(testK*(-1/testGamma));

trainKFinal = [(1:size(trainD, 1))' trainK];
testKFinal = [(1:size(testD, 1))' testK];
end
```

5)

```
optimization finished, #iter = 442
nu = 0.000052
obj = -102.607261, rho = -0.148167
nSV = 177, nBSV = 0
Total nSV = 1050
Cross Validation Accuracy = 93.5847%
   93.5847
```

6)

Name	Submitted	Wait time	Execution time	Score
hw5_1.csv	10 hours ago	0 seconds	0 seconds	0.80375

Complete

Jump to your position on the leaderboard

7)

44	new	**Astitv Nagpal**	0.80375	6	10h

Your Best Entry

Your submission scored 0.80375, which is not an improvement of your best score. Keep trying!