Cover page for answers.pdf

CSE512 Fall 2019 - Machine Learning - Homework 5

Your Name: Shubham Bagi

Solar ID: 112672171

NetID email address: Shubham.bagi@stonybrook.edu

Names of people whom you discussed the homework with:

GuruSangama Prasad.

1. Question 1 - Boosting

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

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

Case I:

Let $y^j = 1$ and $f(x^j) = 1$

$$y^j(f(x)) = 1$$

Case II

Let $y^j = -1$ and $f(x^j) = 1$

$$y^j(f(x)) = -1$$

Case III

let $y^j = 1$ and $f(x^j) = -1$

$$y^j(f(x)) = -1$$

Case IV

let $y^j = -1$ and $f(x^j) = -1$

$$y^j f(x) = 1$$

If $y^j f(x) = 1 \Rightarrow$ error $= 0$

and $y^j f(x) = -1 \Rightarrow$ Here the error $= 1$

We have

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

There exists some z for which $e^{-z} \geq 1$ if $z \leq 0$

$$\therefore e^{(-y^j(f(x^j)))} \geq 1, \text{ if } y^j f(x^j) \leq 0$$

1.2. The weight for each data-point j at step $t+1$ can be defined by:

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

Here Z_t is Normalising Constant

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

$w_j = 1/N$ by weight update rule

$$w_j^2 = \frac{1}{N} e^{\frac{-\alpha_i y_j h_t(x_j)}{Z_1}}$$

$$w_j^3 = \frac{1}{N} e \frac{(-\alpha_i y_j h_i(x_j)) \, e^{(-\alpha_2 y_j h_2(x_j))}}{Z_1 Z_2}$$

We write

$$w_j^{t+1} = \frac{1}{N} e^{\frac{-y_j f(x_j)}{\prod Z_t}}$$

Also, we know the weights of $(t+1) = 1$

$$\Rightarrow \frac{1}{N \prod Z} \sum_{i=1}^{N} e^{-y_j f(x_j)} = 1$$

$$\Rightarrow \frac{1}{N} \sum_{i=1}^{N} e^{-y_j f(x_j))} = \prod_{t=1}^{N} Z_t$$

//

3. a). $\varepsilon_t = \sum_{i=1}^{N} w_i^{(t)} \, \delta(h_t(x^i) \neq y^i)$

be the weighted training error.

The formula for z_t is given by

$$z_t = (1-\varepsilon_t)\exp(-\alpha_t) + \varepsilon_t \exp(\alpha_t).$$

~~Now~~ After differentiating

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

Here $\frac{\partial z_t}{\partial \alpha_t} = 0$

$$\alpha_t = \log_e \sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}} \;//.$$

$$z_t^{opt} = (1-\varepsilon_t)\, e^{-\log_e \sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}} + \varepsilon_t\, e^{\log_e \sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}}$$

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

Now $z_t^{opt} = 2\sqrt{\varepsilon_t(1-\varepsilon_t)} \;//.$

(b). Now we know from part (a)

$$z_t^{opt} = 2\sqrt{\varepsilon_t(1-\varepsilon_t)}.$$

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

The bound

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

when $x = 4\gamma_t^2$, where $0 \leq \gamma_t \leq 1/2$.

$$\ln(1-\gamma\gamma_t^2) \leq -\gamma\gamma_t^2$$

$$\ln 2\sqrt{\frac{1}{\gamma} - \gamma_t^2} \leq -2\gamma_t^2$$

$$\Rightarrow z_t \leq e^{-2\gamma_t^2}$$

(c).

$$\varepsilon_{training} \leq \prod_{t=1}^{T} z_t \leq \exp(-2\sum_{t=1}^{T}\gamma_t^2)$$

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

Now for T to be infimity

$e^{(-2T\gamma_t^2)}$ will tend to zero

So when this term is zero, we will achieve a zero error when the steps is Large

2. Programming question

2.5

1. **For K = [2]**

    I. Iterations= [11]
    II. SumOfSquares = [535871217.752638]
    III. p1= [38.040685]
    IV. p2= [88.010588]]
    V. p3= [63.146431]

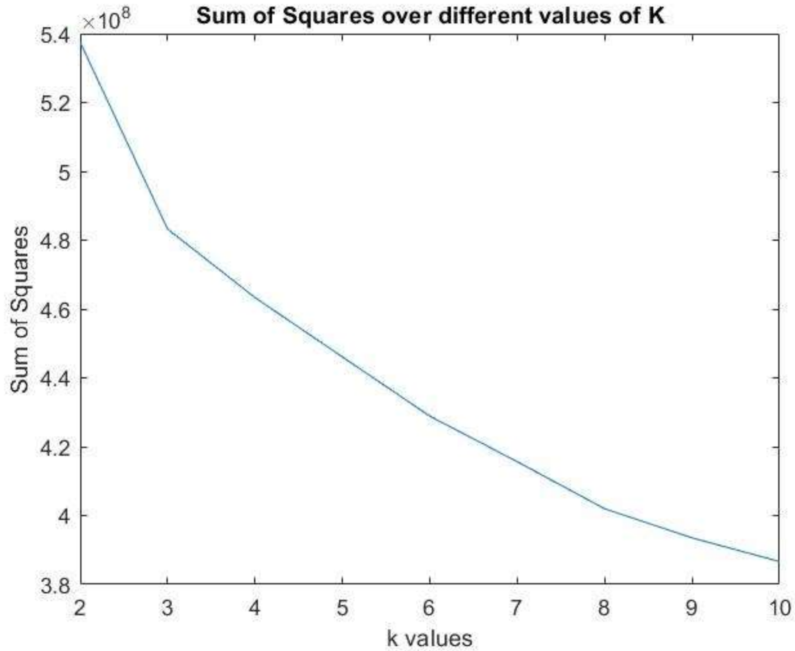
2. **For K = [4]**

    I. Iterations= [10]
    II. SumOfSquares = [461163553.199967]
    III. p1= [63.561606]
    IV. p2= [88.167888]
    V. p3= [75.864747]

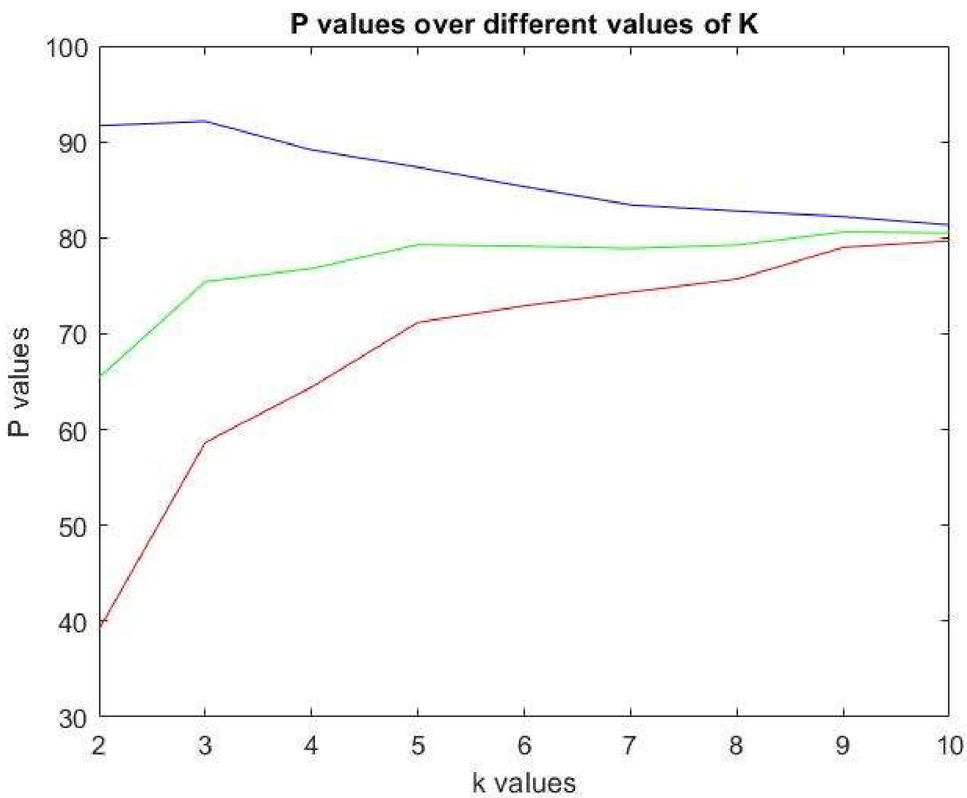
3. **For K = [6]**
    I. Iterations= [13]
    II. SumOfSquares = [422830896.339120]
    III. p1= [66.702273]
    IV. p2= [82.985707]
    V. p3= [74.843990]

2. Number of iterations for K = 6 is **Iterations = 13.**

3. Plot of Sum of squares vs K.



4. Plot of P1,P2,P3 for different values of K.



3. Programming Question

3.4.2

For default value of C and Gamma.

Cross Validation Accuracy = 15.6443%

3.4.3

After tuning

Cross Validation Accuracy = 58.5256% when s = 0, v =5, c = 10, g =1

Cross Validation Accuracy = 73.7198% when s = 0, v = 5, c = 20, g = 5

Cross Validation Accuracy = 77.4902% when s = 0, v = 5, c = 20, g = 10

Cross Validation Accuracy = 81.0355% when s = 0, v = 5, c = 50, g = 10

Cross Validation Accuracy = 82.4423% when s = 0, v = 5, c = 100, g = 10

Cross Validation Accuracy = 84.018% when s=0, v=5, c =100, g=20

Cross Validation Accuracy = 84.1306% when s =0, v =5, c = 150, g = 20

Cross Validation Accuracy = 84.1306% when s =0, v =5, c = 1000, g = 2

Cross Validation Accuracy = 85.4249% when s =0, v =5, c =2000, g = 2

When Clusters are increased the C V Accuracy is decreasing.

So best value of C and Gamma which yields better Accuracy is :

C = 2000 and G = 2

3.4.5

Using Chi-square Method to determine the accuracy of the model.

For default value of gamma and C =1

Cross Validation Accuracy = 44.6258%

Cross Validation Accuracy = 74.1137% when s =0, v =5, c =10, g = 1

Cross Validation Accuracy = 85.8188% when s =0, v =5, c = 100, g =1

Cross Validation Accuracy = 87.1694% when s =0, v =5, c = 200, g =1

Cross Validation Accuracy = 89.1953% when s =0, v =5, c = 500, g =1

Cross Validation Accuracy 89.0827 for higher values of C and when g =1

So the best value of accuracy is got when **C value is 500 and g value is 1**.

Kaggle Score – 0.71666

The Accuracy got is 71.66%