CS 6375

ASSIGNMENT \_1\_\_\_\_\_\_\_\_\_\_\_

Names of students in your group: Yiming Zhou

Number of free late days used: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

1.

|  |  |  |  |
| --- | --- | --- | --- |
| Hyperparameter: | | α= 0.25 |  |
| Steps |  |  | Error |
| 0 | 0 | 1 | 1 |
| 1 | 0 | 0.625 | 0.7890625 |
| 2 | 0.1875 | 0.859375 | 0.6363525390625 |
| 3 | 0.2109375 | 0.61914062 | 0.5234165191650391 |
| 4 | 0.34863281 | 0.75756836 | 0.4381965696811676 |
| 5 | 0.38269043 | 0.60220337 | 0.3727042549289763 |

The table shown above are steps of gradient descent. I choose learning rate to be 0.25. The training error goes down after each iteration.

2.

False Positive Rate = 0.1

False Negative Rate = 0.2

3.

The most specific hypothesis S best fits positive examples in the training data and the most general hypothesis G fits the negative examples well and generalize all the hypotheses that are more general to S. Given that training data has no errors and large enough so that it can cover most of instance space, the resulting hypothesis will converge correctly.

However, since the most specific hypothesis and the most general hypothesis can only learn from training data, they could not ‘clean’ noisy data but being mislead by those data.

4.

Consistent: A hypothesis h is consistent with a set of training examples D if and only if h(x)=c(x) for each example <x, c(x)> in D.

Version Space: The version space is the subset of hypotheses from H consistent with the training examples in D.

5.

all possible

6.

(a)2\*2\*2\*2 = 16

(b)3\*3\*3\*3 = 81

(c)3\*3\*3\*3\*2 = 162

(d)6\*4=24

(e)6\*4\*2=48

7.

|  |  |  |
| --- | --- | --- |
| Step | Example | S Hypothesis |
| 0 | / | <(Ф, Ф, Ф, Ф),( Ф, Ф, Ф, Ф)> |
| 1 | (<1,1,0,1,1>,1) | <1,1,0,1,1> |
| 2 | (<0,1,0,1,1>,0) | <1,1,0,1,1> |
| 3 | (<1,1,1,1,0>,1) | <1,1,?,1,?> |
| 4 | (<0,0,0,1,1>,0) | <1,1,?,1,?> |
| 5 | (<1,1,1,1,1>,1) | <1,1,?,1,?> |

8.

9.

(a)

S boundary: (x1=4, x2=6, y1=3, y2=5)

(b)

G boundary: (x1=3, x2= 8, y1=2, y2=7)

(c)

Query (x=4, y=6) with label positive can reduce the version space. And query (x=5, y=4) with label positive will not reduce the version space.

(d)

The smallest number of training examples to learn the target concept is 2. They could be (x=3,y=2) with label positive and (x=5,y=9) with label positive.

10.

(a)

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Input example | G hypothesis | S hypothesis |
| 1 | / | <(?,?,?,?),(?,?,?,?)> | <(Ф, Ф, Ф, Ф),( Ф, Ф, Ф, Ф)> |
| 2 | <(ug,se,l,hs),(gr,cs,h,hs)>,+ | <(?,?,?,?),(?,?,?,?)> | <(ug,se,l,hs),(gr,cs,h,hs)> |
| 3 | <(ug,se,h,fr),(gr,cs,h,hs)>,+ | <(?,?,?,?),(?,?,?,?)> | <(ug,se,?,?),(gr,cs,h,hs)> |
| 4 | <(gr,se,l,so),(gr,cs,h,se)>,- | {<(ug,?,?,?),(?,?,?,?)>,  <(?,?,?,?),(?,?,?,hs)>} | <(ug,se,?,?),(gr,cs,h,hs)> |
| 5 | <(ug,se,l,ju),(gr,se,h,ju)>,+ | <(ug,?,?,?),(?,?,?,?)> | <(ug,se,?,?),(gr,?,h,?)> |

(b)

There are total (1\*2\*1\*1)\*(2\*1\*2\*1)=8 consistent hypotheses. There are only (1\*1\*1\*1)\*(1\*1\*1\*1)=1 consistent with example <(ug,cs,h,do),(gr,ma,l,se)> with class label “+”.