In [1]:
```
import math
import numpy as np
```

In [2]:
```
f = open("pa5train.txt", "r")
train_data = [[int(i) for i in line.strip().split()] for line in f]

f1 = open("pa5test.txt", "r")
test_data = [[int(i) for i in line.strip().split()] for line in f1]
```

In [3]:
```
def h( data, word, sign):
    list1 = [sign*(1 if line[word] else -1) for line in data]
    return list1

def err( data, H , D):
    err = np.dot(D,[H[i] != data[i][-1] for i in range(len(H))])
    return err

def alpha(err):
    return 1/2*math.log((1-err)/err)
```

In [4]:
```
def boost(data, t):
    classifer = []
    D  = [1/len(data)]*len(data)

    for i in range(t):
        err_list = [(err(data, h(data, i ,sign), D),i,sign) for i in ran
ge(4003) for sign in [1,-1]]
        error,word,sign = min(err_list)
        if (error >= 0.5):
            break

        ht = h(data, word, sign)
        al = alpha(error)
        D = [D[i]*math.exp(-al*data[i][-1]*ht[i]) for i in range(len(D
))]
        Z = sum(D)
        D = [D[i]/Z for i in range(len(D))]
        classifer += [(al,word,sign)]
    return classifer
```

In [5]:
```
def error(data, classifier):
    count = 0
    for line in data:
        sum = 0
        for (alpha, word, sign) in classifier:
            h = 1*sign if line[word] else -1*sign
            sum += alpha*h
        count += 1 if sum*line[-1] < 0 else 0
    return count/len(data)
```

Q1

In [6]:
```
for i in [4,3,7,10,15,20]:
    classifier = boost(train_data,i)
    print("t =",i)
    print("train_error:",error(train_data, classifier))
    print("test_error:",error(test_data, classifier))
```

```
t = 4
train_error: 0.051111111111111114
test_error: 0.03875968992248062
t = 3
train_error: 0.06444444444444444
test_error: 0.03875968992248062
t = 7
train_error: 0.028888888888888888
test_error: 0.031007751937984496
t = 10
train_error: 0.015555555555555555
test_error: 0.03875968992248062
t = 15
train_error: 0.0
test_error: 0.023255813953488372
t = 20
train_error: 0.0
test_error: 0.023255813953488372
```

Q2

In [7]:
```
f = open("pa5dictionary.txt", "r")
dictionary = [w.strip() for w in f]
```

In [8]:
```
classifier = boost(train_data,10)
words = [dictionary[w] for (e,w,s) in classifier]
print(words)
```

```
['remove', 'language', 'free', 'university', 'money', 'linguistic', 'cl
ick', 'fax', 'want', 'de']
```