**Assignment 1 Part 1 Report for CSE-512**

1. **A warm-up problem**

Ans. According to the hypothesis h defined in the problem, there can be 2 errors,



Training error is defined as the number of wrong predictions divided by the total predictions.

Hence in our case,

Probability of correct predictions = 1 – = 1 – 0.1 = 0.9

1. **Bayes Optimal Predictor**

Ans. We have to prove **--- 1** and we have been given classifier with

Loss function is the probability of output of our function does not match the value of y i.e.,

=

By Law of total probability:

Taking the minimum value of 1 and 1-t to minimize the error, we get:

Above is the Bayes optimal predictor, thus we have proved the **inequality 1**.

1. **Unidentical distributions**

Suppose we have a hypothesis h such that

As the sequence is independent, we can take one sample each from each distribution.

We know that,

We get above from the fact that which is the arithmetic-geometric inequality.

From the union bound , we can say that the probability exists where ℇ such that and .

Thus,

1. **Vapnik-Chervonenkis (VC) dimension**
2. VC dimension is defined as the capacity of space of functions that could be learned by the classification algorithm or the cardinality of largest number of points which could be shattered by the learning algorithm.

Let’s consider a k such that k(x) = 1 if x is inside the rectangle otherwise k(x) = 0, where . As the points would be covered by a rectangle atleast, the VC dimension would be 4 at the minimum. Suppose some points come at the axes the labelling cannot be done for them as they aren’t in the boundaries of the rectangle. For example, we take 5 points, 4 are on axes 1 on origin, we would not be able to label the points on the axes.

Suppose we take real numbers such that

A classifier such that

and

Let’s take where

or

We observe that the above gets shattered hence VC is 2d.

Considering example if and otherwise.

if and otherwise.

Then for

If K is a set of size (), as per pigeonhole principle, we can say that exists in K such that,

For every in |d| and exists in K when . Now the points are 2d+1 then minimum 1 point must lie inside the rectangle. And if we label this point as negative then this labelling cannot be separated by any rectangle. This proves .

If we consider positive labelling points, we can say that,

Therefore, it is proved that

We have been given hypothesis class of sine functions: where , [Replacing x with its binary representation ]

Multiplying with shift bits

When we take = 0, the equation is positive, since the sine function is positive in Quadrant 1 and 2, .

When we take = 1, then the above equation is negative, since the sine function is negative in Quadrant 3 and 4 and

Taking points which can be shattered by H i.e. and ranges from and all the values of

Now, we can get the 1st bit column in binary expression using

Similarly, we can get the 2nd but column in binary expansion using

As we keep repeating the above steps taking increasing values of n we would see that it gets shattered. Hence, we prove that

1. **Boosting**

So, the error on the iteration’s will be and the error of the weak learner on the next iteration’s

Also,

Also,

So, now dividing the above two equations:

Therefore,

1. **Learnability of logistic regression**

We have a loss function

Taking a function defined by *g(a) = log (1 + )*

If is for some .

f is convex if g is convex, hence

We have to prove that g(a) is learnable by lipschitzness(convex lipschitz bounded) and smootheness (convex smooth bounded) in order to show that a function is learnable.

Lipschitzness:

As *g(a) = log (1 +* ), taking a derivative of this equation we get

Above is 1 – Lipschitz.

Let where, is -Lipschitz and is -Lipschitz.

Then we know that f is -Lipschitz.

If is a linear function, then for some   
  
This implies, f is - Lipschitz

Hence, is B-Lipschitz.

Smoothness:

Double derivative of *g(a)* is given by

We have, where is a -smooth function , , then f is -smooth.

Using above and mean value theorem gives is 1/4 Lipschitz. Thus, is – smooth.

Hence parameters for Convex-smooth bounded and Convex-lipschitz-bounded are /4 and B respectively.

1. **Learnability of Halfspaces with hinge loss**

We consider , and

Let for .

To prove that , considering

We have

We assume that

Assuming without any loss of generality,

1. **Cross validation**

Given the labels (0,1) are chosen randomly with probability 0.5

For an i.i.d training sample S with output hypothesis as h, the true error is 0.5 since h is a constant function labeling based on the parity values of labels.

Thus, the error will be defined by the total number of incorrect predictions/total number of predictions. Thus,

For parity of S =1 and fold set(x,y) in S,

Considering the parity of S/x is 1, then y = 0. The constant prediction output h(x) of training using S/x will be 1 and the LOO using this fold is 1.

Considering the parity of S/x is 0, then y = 1. The constant prediction output h(x) of training using S/x will be 0 and the LOO using this fold is again 1.

The estimate of error of h is 1 when we calculate the average over folds.

Similarly, for parity of S=0, the estimate of error of h averaging over folds is 1.

Evaluating the difference between error and estimate we get –

1. **Local minimum**

We consider H as a class of homogenous Halfspaces in

Let, sample ,

Then and

W is a local minima, we take .

By Cauchy–Schwarz inequality we have, for every w’with

Thus, we obtain