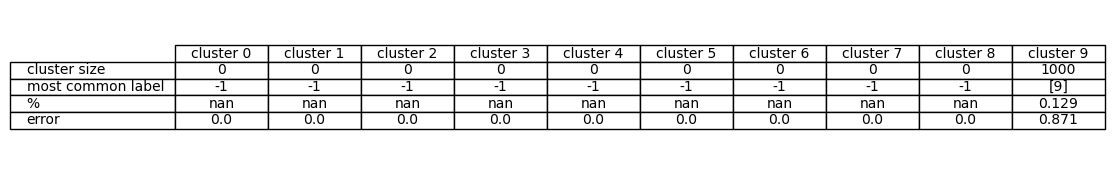
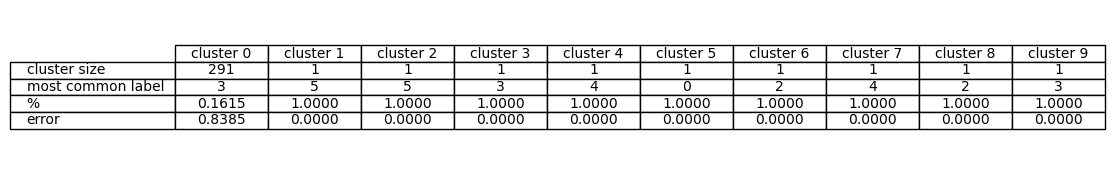
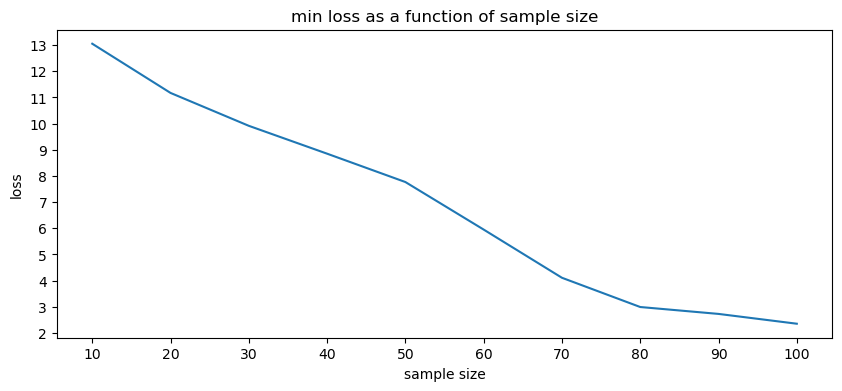
1.

(c)

(d)

2.(a)

(b)

As the training set grows, we expected the loss on the distribution to reduce. Based on what we’ve seen in class: . we used the that minimizes and as grows, gets smaller.

(c) yes.

(d)

3.(a)

According to GD algorithm:

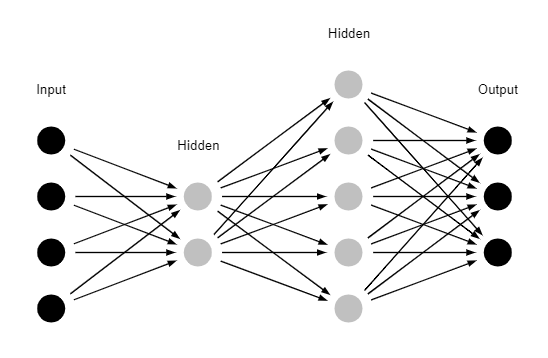
Let

we will find

=

Hence,

(b)

4.(a)

(b).

Since the input layer has 4 neurons,

(c)

Since ,

(d)

Then

5.(a)

For each sample there are attributes, and we can test 3 options for , hence there are options for inner vertex (+2 for label). In a tree of depth there are at most nodes, thus

(b)

Danny is wrong, since ID3 is heuristic algorithm and not ERM it does not guarantee minimal error. After performing pruning, we can’t guarantee we will get minimal error. Thus PAC assumption on the sample size does not apply.

6.(a)

The assumption **does not hold**.

Naïve bayes assumption is

But in the given distribution,

and

And

Which means for some

(b)

From previous section, we know we can’t use the naïve bayes assumption.

Hence

For we get:

for the predictor will return

For we get:

for the predictor will return

For we get:

for the predictor will return

For we get:

for the predictor will return

Therefore ,

7.(a)

From the definition of we can see that there’s a linear dependence between the first and second coordinates to the third and fourth coordinates, hence the degree of is 2 and therefore the degree of is 2 as well. Since and the degree of is 2, therefore it has 2 eigenvalues 0.

it has kernel of dimension 2, and from the dimension theorem (

we know that the eigenvalues of are non-negative. thus, the 2 smallest eigenvalues of are both 0 and the distortion is

(b)

Let ,

Note that all satisfy

For

After performing gaussian elimination on we get

We got all the eigenvalues are 1, therefore the distortion is sum of positives and .