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)

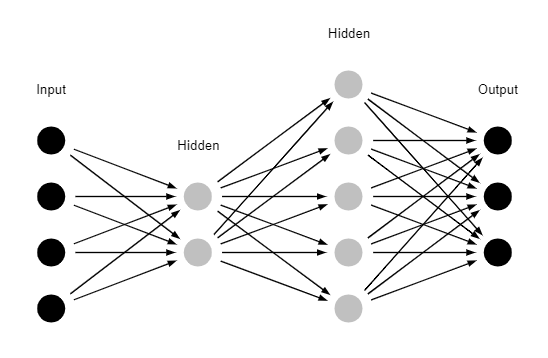
3.(a)

According to GD algorithm:

Let

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 tests 3 options for , hence there are options for inner vertex (+2 for label). In a tree of depth there are at most znodes, thus

(b)

Danny is wrong, since ID3 is not ERM algorithm we can’t guarantee such thing.

6.(a)

The assumption does not hold.

Naïve bayes assumption is

But in the given distribution,

and

And

Which means for some

(b)

, ,

Hence ,

So the predictor we got is

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 (