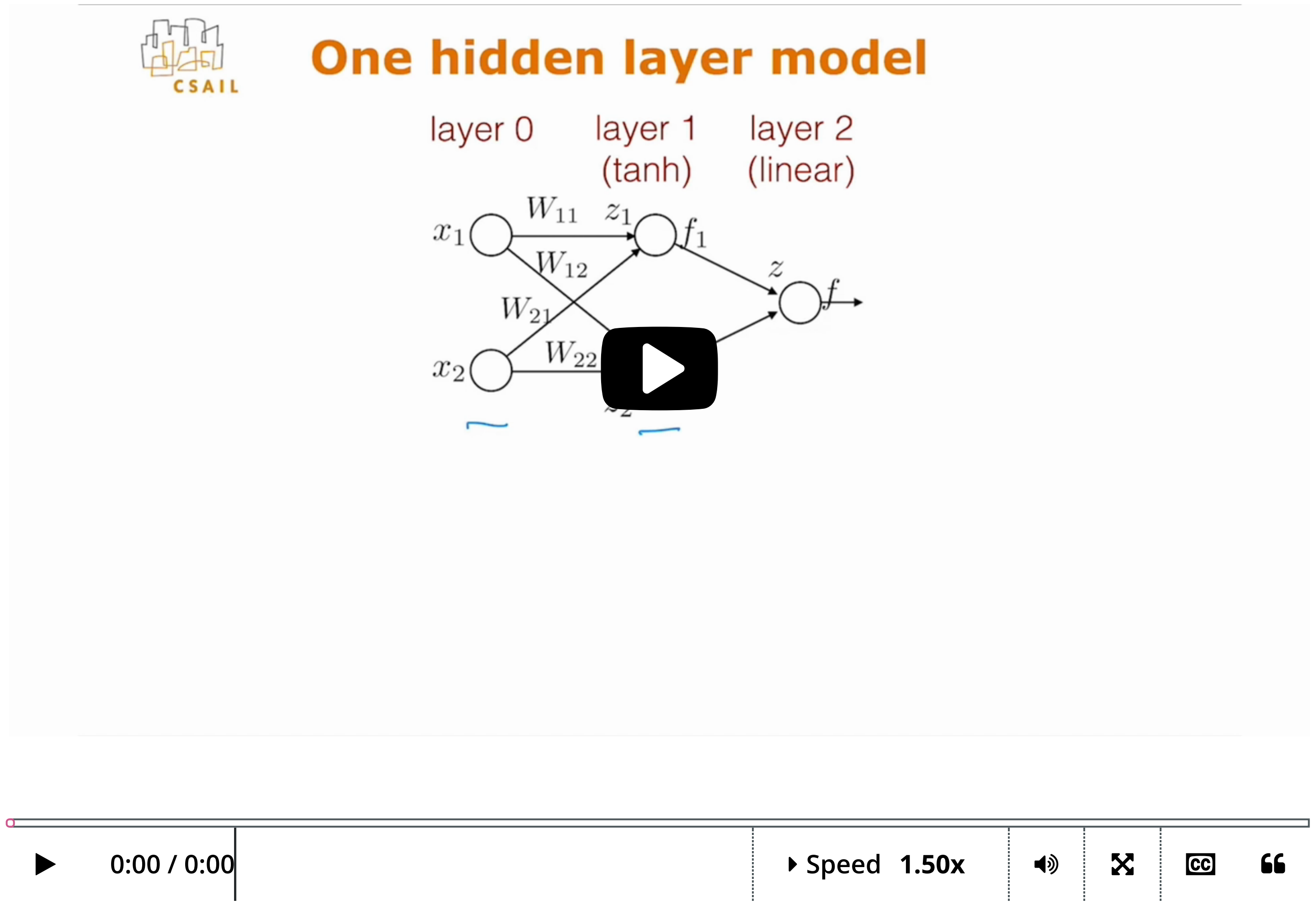
MITx: 6.86x
Machine Learning with Python-From Linear Models to Deep Learning

Help sandipan_dey

Course > Unit 3 Neural networks (2.5 weeks) > Lecture 8. Introduction to Feedforward Neural Networks > 6. Hidden Layer Models

6. Hidden Layer Models

Models with Hidden Layer



One hidden layer model

layer 0 layer 1 (tanh) layer 2 (linear)

Start of transcript. Skip to the end.

OK.

Let's look at these models in a little bit more detail now,

trying to understand them, how the computation is performed,

and how to visualize what they can and cannot do.

So here I have a simple three-layer neural network

feed forward with the input layer, one hidden layer,

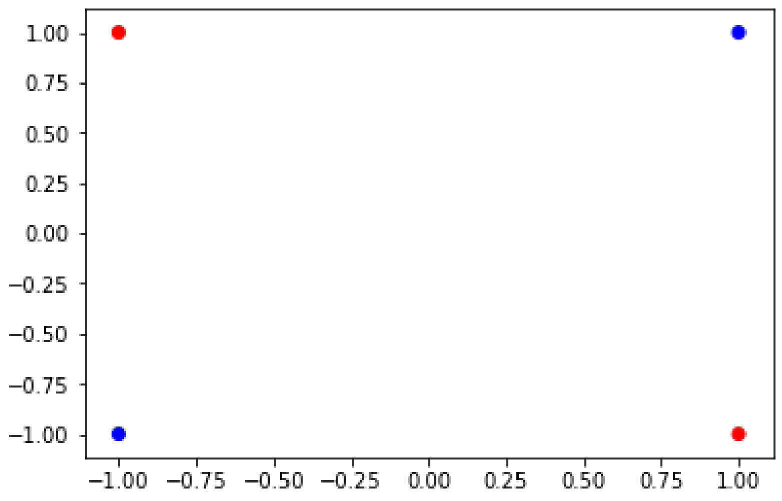
Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

For the following set of problems, let's consider a simple 2-dimensional classification task. The training set is made up of 4 points listed below:

$$\begin{aligned} x^{(1)} &= (-1,-1) \quad , \quad & y^{(1)} &= 1 \\ x^{(2)} &= (1,-1) \quad , \quad & y^{(2)} &= -1 \\ x^{(3)} &= (-1,1) \quad , \quad & y^{(3)} &= -1 \\ x^{(4)} &= (1,1) \quad , \quad & y^{(4)} &= 1 \end{aligned}$$

The dataset is illustrated below (blue - positive, red - negative)

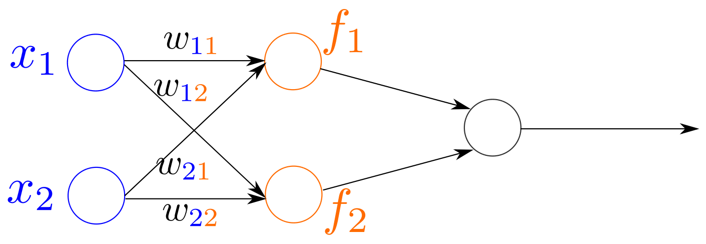


For simplicity, assume that we are only interested in binary classification problems for now. That is, $y^{(i)}$ can be either 1 or -1.

Linear Separability After First Layer

1/1 point (graded)

For this problem, let us focus on a network with one hidden layer and two units in that layer:



Let $f_1^{(i)}, f_2^{(i)}$ denote the output of the two units in the hidden layer corresponding to the input $x^{(i)}$ respectively, i.e.

$$\begin{aligned} f_1^{(i)} &= f\left(w_{01} + \left(w_{11}x_1^{(i)} + w_{21}x_2^{(i)}\right)\right) \\ f_2^{(i)} &= f\left(w_{02} + \left(w_{12}x_1^{(i)} + w_{22}x_2^{(i)}\right)\right) \end{aligned}$$

Consider the set $D' = \left\{\left(\left[f_1^{(i)}, f_2^{(i)}\right], y^{(i)}\right), \quad i = 1,2,3,4\right\}$.

Assume that f is the linear activation function given by $f(z) = 2z - 3$.

For which of the following values of weights would the set D' be linearly separable? (Select all that apply.)

☐ $w_{11} = w_{21} = 0, w_{12} = w_{22} = 0, w_{01} = w_{02} = 0$

☐ $w_{11} = w_{21} = 2, w_{12} = w_{22} = -2, w_{01} = w_{02} = 1$

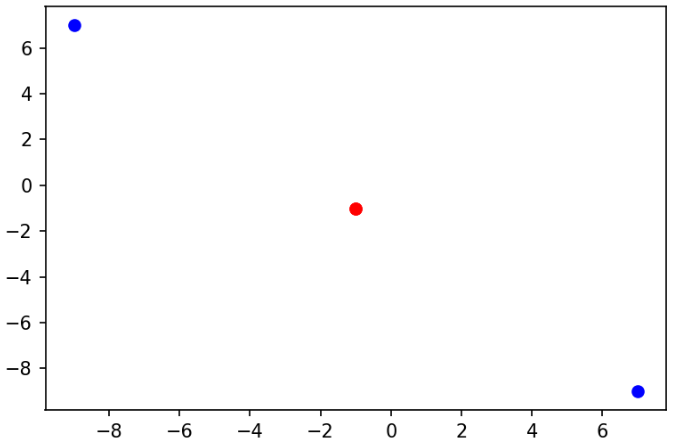
☐ $w_{11} = w_{21} = -2, w_{12} = w_{22} = 2, w_{01} = w_{02} = 1$

☑ None of the above ✓

✓

Solution:

First of all note that from the figure in the text above that D is clearly not linearly separable.
Also, $f(z) = 2z - 3$ is a linear activation function.
Any linear transformation of the feature space of a linearly in-separable classification problem would still continue to remain linearly inseparable. For this question, one can compute the feature representations of all the data points and verify visually.
For example, the result of the second answer is plotted here:



Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Non-linear Activation Functions

1/1 point (graded)

Again, let's focus on a network with one hidden layer with two units and use the same training set as above. The weights of the network are given as follows:

$$\begin{aligned} w_{11} &= 1, w_{21} = -1, w_{01} = 1 \\ w_{12} &= -1, w_{22} = 1, w_{02} = 1 \end{aligned}$$

Let f_1, f_2 be the outputs of the first and second unit respectively.

Consider the set $D' = \{([f_1^{(i)}, f_2^{(i)}], y^{(i)}), \quad i = 1, 2, 3, 4\}$

For which of the following functions f, would the set D' be linearly separable? (Select one or more that apply.)

□ $f(z) = 5z - 2$

☑ $f(z) = \mathrm{ReLU}(z)$ ✓

☑ $f(z) = \tanh(z)$ ✓

□ $f(z) = z$

✓

Solution:

From the above problem, we note that any linear transformation of the feature space of a linearly in-separable classification problem would still continue to remain linearly inseparable. Hence we rule out the two linear functions.

For all of the parts below, note that

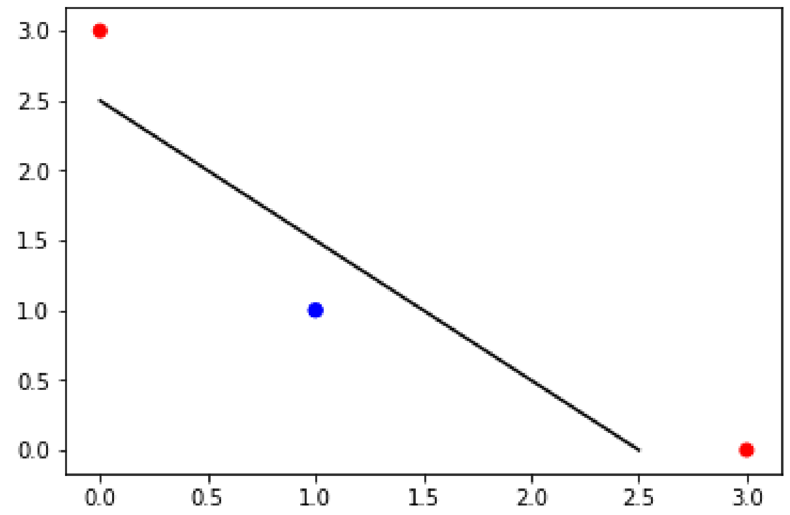
$$f_1^{(i)} \quad = \quad f(w_{01} + (w_{11} x_1^{(i)} + w_{21} x_2^{(i)}))$$

$$f_2^{(i)} \;=\; f\left(w_{02} + \left(w_{12}x_1^{(i)} + w_{22}x_2^{(i)}\right)\right)$$

- $f\left(z\right) = \mathrm{ReLU}\left(z\right)$: Substituting for ReLU into f in the above equation gives the following results:

$$\left(1,1\right), \left(3,0\right), \left(0,3\right), \left(1,1\right)$$

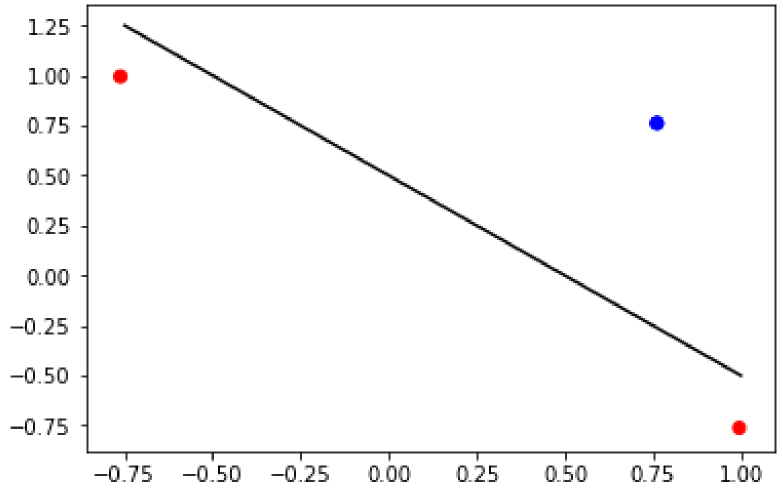
The following figure plots these points and a potential linear classifier:



- $f\left(x\right) = \tanh\left(x\right)$: Substituting for $tanh$ into f in the above equation gives the following results:

$$\left(0.76, 0.76\right), \left(0.99, -0.76\right), \left(-0.76, 0.99\right), \left(0.76, 0.76\right)$$

The following figure plots these points and a potential linear classifier:



Submit

You have used 2 of 2 attempts

Answers are displayed within the problem

Neural Network Learned parameters

1/1 point (graded)

Given a neural network structure and an objective function, the parameters that are optimized (i.e. learned) during the training process are:

☐ The dimension of the feature representation

☑ The parameters that control the feature representation ✓

☐ The hyper-parameters

☑ The parameters for the classifier ✓

✓

Solution:

Similar to the linear classifiers that we covered in previous lectures, we need to learn the parameters for the classifier. However, in this case we also learn the parameters that generate a representation for the data.

The dimensions and the hyper-parameters are decided with the structure of the model and are not optimized directly during the learning process but can be chosen by performing a grid search with the evaluation data or by more advanced techniques (such as meta-learning).

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 3 Neural networks (2.5 weeks):Lecture 8. Introduction to Feedforward Neural Networks / 6. Hidden Layer Models

Add a Post

Show all posts ▼ by recent activity ▼

? Non-linear Activation Functions (staff)
Why the answer is this? It is very hard to write without giving out the answer. So, I will do. Please edit latter. In every case, we end up --- edited --- Obviously, I understand what... 6

? 10 random hidden units trained?
I thought the point of the last example was to show that if you randomly choose a sufficiently large number of hidden layers you would find some that would make the points... 4
Community TA

? [Staff] Neural Network Learned parameters
I don't quite understand options for this question, and what answer is expected: 1. What are "The parameters that control the feature representation" ? 2. I assume that "The ... 6

? Should we use programming to solve first 2 questions?
Should we use programming to solve first 2 questions? Or should we only solve by hand, I find it a bit complicated solving by hand. 3

? Linear Separability After First Layer Exercise - Incorrect weight (w) notation in the figure
Maybe there is an incorrect weight (w) notation in the figure? The formula is OK. 2

Learn About Verified Certificates