Neural Networks Solution

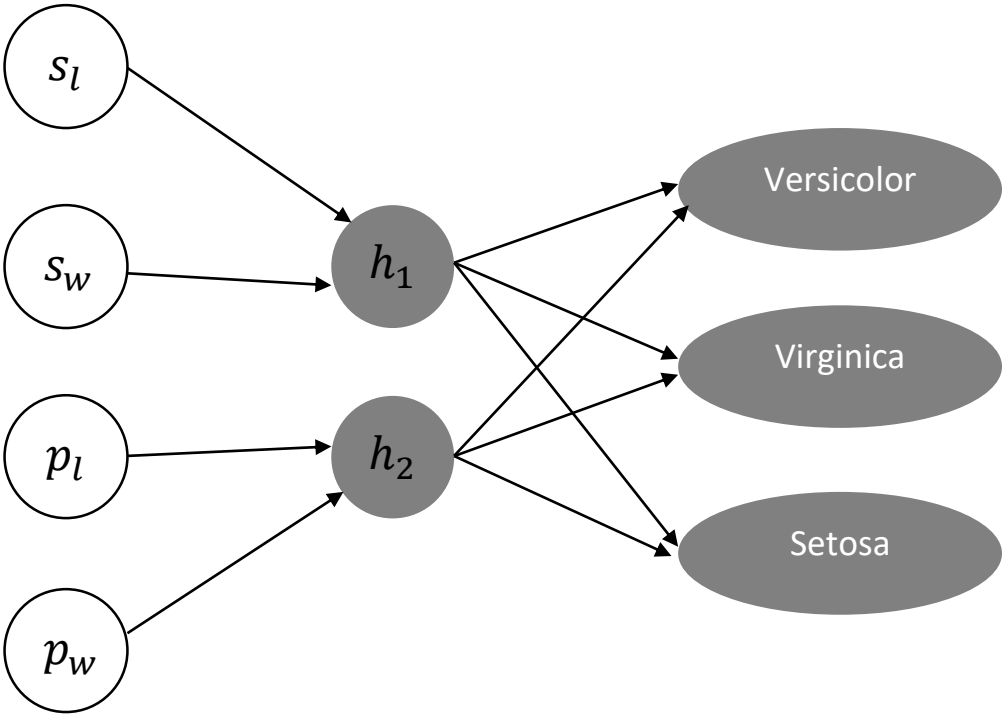
Fisher's "iris" dataset was introduced in 1936 and serves as a common tool for practitioners to practice or test different classification techniques. It contains 150 observations along four measurements for three species of Iris flowers, with 50 observations each species. For these problems, consider the following subset of the famous "iris" dataset in Table 1.

Table 1. Attributes of flowers and their species, subsetted from the "iris" dataset

Sepal.Length s_l	Sepal.Width s_w	Petal.Length p_l	Petal.Width p_w	Species s
6.7	2.5	5.8	1.8	virginica
6.1	3	4.6	1.4	versicolor
4.9	3	1.4	0.2	setosa
5.4	3.7	1.5	0.2	setosa
5.8	2.7	4.1	1	versicolor
6.1	2.6	5.6	1.4	virginica
4.7	3.2	1.3	0.2	setosa
5.1	3.8	1.6	0.2	setosa
6.3	2.5	5	1.9	virginica
4.4	3.2	1.3	0.2	setosa
5	3.4	1.6	0.4	setosa
6.5	3	5.2	2	virginica
4.6	3.6	1	0.2	setosa
4.8	3	1.4	0.1	setosa
6.3	2.8	5.1	1.5	virginica
6.3	3.3	4.7	1.6	versicolor
6.1	2.9	4.7	1.4	versicolor
5.4	3	4.5	1.5	versicolor
7.7	2.8	6.7	2	virginica
5.2	3.4	1.4	0.2	setosa

Say we construct a neural network to classify the observations based on the four variables (Figure 1) with a generic activation function φ. The predicted class is determined by finding the biggest activation in the final column of neurons.

Figure 1. Neural network for “iris” dataset



Note: Any gray node has the following activation function:

$$\varphi(x) = \frac{1}{1 + e^{-x}}$$

Table 2. Weights on network

The weight is	from	to
0.2	s_l	h_1
0.4	s_w	h_1
0.6	p_l	h_2
0.3	p_w	h_2
0.5	h_1	Versicolor
0.3	h_1	Virginica
0.6	h_1	Setosa
0.1	h_2	Versicolor
0.7	h_2	Virginica
0.2	h_2	Setosa

Questions

1. What is the performance of a classifier that randomly guessed the class of each observation?

   Answer: 33%.
   Since there are three possible classes, random assignment would give us a 1 in 3 chance of guessing the correct class. **So the classifier's accuracy is ~33%.** The fact the sizes of the classes are different is irrelevant.

2. Evaluate the performance of the network with all weights equal to 1.

   a) the network predicts only a single class

   b) the network cannot predict any class with confidence

   c) the network predicts all observations perfectly, but fails given new data

   d) the network only is accurate with 50 percent of the observations

   Answer: B.
   If all the weights are 1, then the classifier cannot make a prediction. All activations in the last layer will be equal. In a sense, it will be guessing at random like in question 1.

3. What is the activation in Versicolor if all the weights are equal to 1 for the first observation?

   Answer: 0.6351

   The inputs and intended class is:

6.7	2.5	5.8	1.8	virginica

   And the relevant weights are:

The weight is	from	to
0.2	s_l	h_1
0.4	s_w	h_1
0.6	p_l	h_2
0.3	p_w	h_2
0.5	h_1	Versicolor
0.1	h_2	Versicolor

This means the activation in the first two neurons are:

$$A(h_1) = \frac{1}{1 + e^{-(0.2*6.7+0.4*2.5)}}$$

$$A(h_2) = \frac{1}{1 + e^{-(0.6*5.8+0.3*1.8)}}$$

Versicolor's activation is then:

$$A(versicolor) = \frac{1}{1 + e^{-(0.5*A(h_1)+0.1*A(h_2))}} = \frac{1}{1 + e^{-\left(\frac{0.5}{1+e^{-(0.2*6.7+0.4*2.5)}}+\frac{0.1}{1+e^{-(0.6*5.8+0.3*1.8)}}\right)}}$$
$$= \mathbf{0.6351}$$