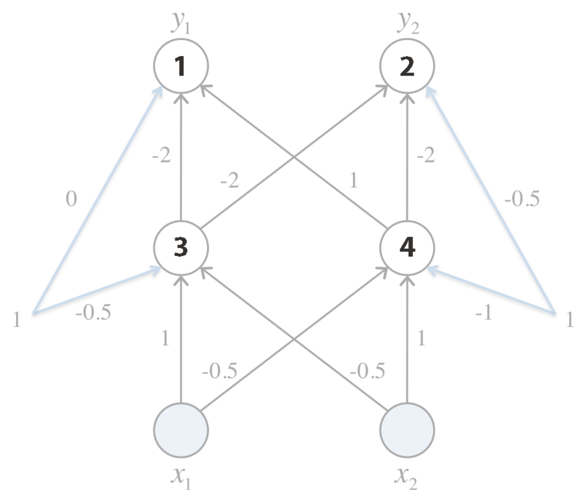
1. In order to calculate the backpropagation of error, we have to follow four steps. I'll number the states as below (for reference from equations).



(a) *Calculate the error of the output units* – start by calculating the output of o_1, o_2, o_3, and o_4.

$$net_3 = 1(x_1) - 0.5(x_2) - 0.5 = 1(0) - 0.5(1) - 0.5 = -1$$
$$o_3 = \frac{1}{1+\exp^{-net_3}} = 0.269$$
$$net_4 = -0.5(x_1) + 1(x_2) - 1 = -0.5(0) + 1(1) - 1 = 0$$
$$o_4 = \frac{1}{1+\exp^{-net_4}} = 0.5$$
$$net_1 = -2(o_3) + 1(o_4) + 0 = -2(0.269) + (0.5) = -0.038$$
$$o_1 = \frac{1}{1+\exp^{-net_1}} = 0.491$$
$$net_2 = -2(o_3) - 2(o_4) - 0.5 = -2(0.269) - 2(0.5) - 0.5 = -2.038$$
$$o_2 = \frac{1}{1+\exp^{-net_2}} = 0.115$$

Now that we have the outputs, the error is as follows:

$$\delta_1 = o_1(1-o_1)(y_1-o_1) = 0.491(1-0.491)(1-0.491) = 0.127$$
$$\delta_2 = o_2(1-o_2)(y_2-o_2) = 0.115(1-0.115)(0-0.115) = -0.012$$

(b) Now, calculate the updates for the output unit weights:

$$\Delta w_{10} = \eta\delta_1 o_0 = 0.1(0.127)(1) = 0.0127$$
$$\Delta w_{13} = \eta\delta_1 o_3 = 0.1(0.127)(0.269) = 0.00342$$
$$\Delta w_{14} = \eta\delta_1 o_4 = 0.1(0.127)(0.5) = 0.00635$$

$$\Delta w_{20} = \eta\delta_2 o_0 = 0.1(-0.012)(1) = -0.0012$$
$$\Delta w_{23} = \eta\delta_2 o_3 = 0.1(-0.012)(0.269) = -0.00032$$
$$\Delta w_{24} = \eta\delta_2 o_4 = 0.1(-0.012)(0.5) = -0.0006$$

(c) Next, calculate the error for the hidden units:

$$\begin{aligned}\delta_3 &= o_3\,(1-o_3)\sum_{\{1,2\}}^{k}\delta_k w_{k3} = o_3\,(1-o_3)\,[\delta_1 w_{13} + \delta_2 w_{23}]\\ &= 0.269(1-0.269)[0.127(-2) - 0.012(-2)] = -0.0452\\ \delta_4 &= o_4\,(1-o_4)\sum_{\{1,2\}}^{k}\delta_k w_{k4} = o_4\,(1-o_4)\,[\delta_1 w_{14} + \delta_2 w_{24}]\\ &= 0.5(1-0.5)[0.127(1) - 0.012(-2)] = 0.0378\end{aligned}$$

(d) Finally, calculate the weight updates for the hidden units:

$$\begin{aligned}\Delta w_{30} &= \eta\delta_3 o_0 = 0.1(-0.0452)(1) = -0.00452\\ \Delta w_{3x_1} &= \eta\delta_3 x_1 = 0.1(-0.0452)(0) = 0\\ \Delta w_{3x_2} &= \eta\delta_3 x_2 = 0.1(-0.0452)(1) = -0.00452\end{aligned}$$

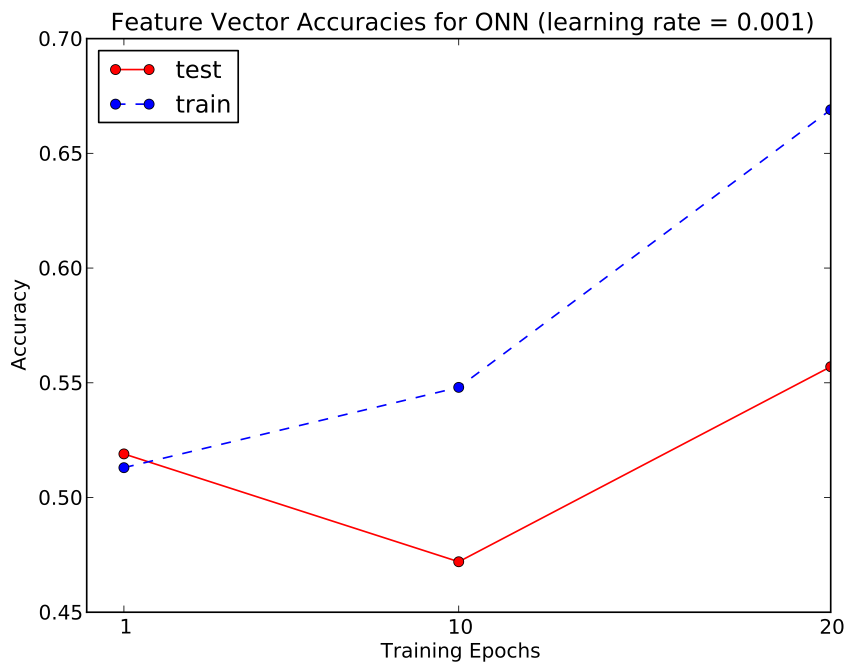
$$\begin{aligned}\Delta w_{40} &= \eta\delta_4 o_0 = 0.1(0.0378)(1) = 0.00378\\ \Delta w_{4x_1} &= \eta\delta_4 x_1 = 0.1(0.0378)(0) = 0\\ \Delta w_{4x_2} &= \eta\delta_4 x_2 = 0.1(0.0378)(1) = 0.00378\end{aligned}$$

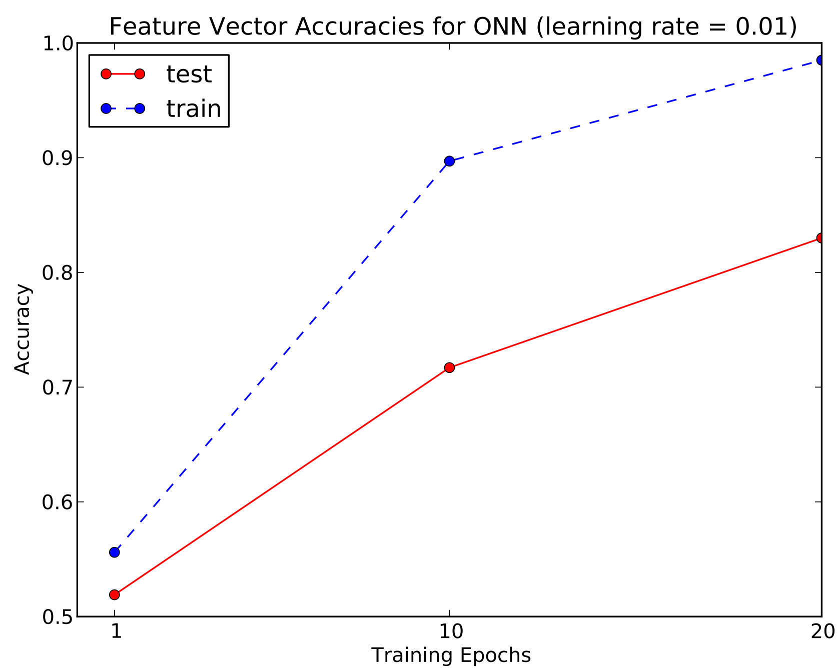
The final weights are therefore:

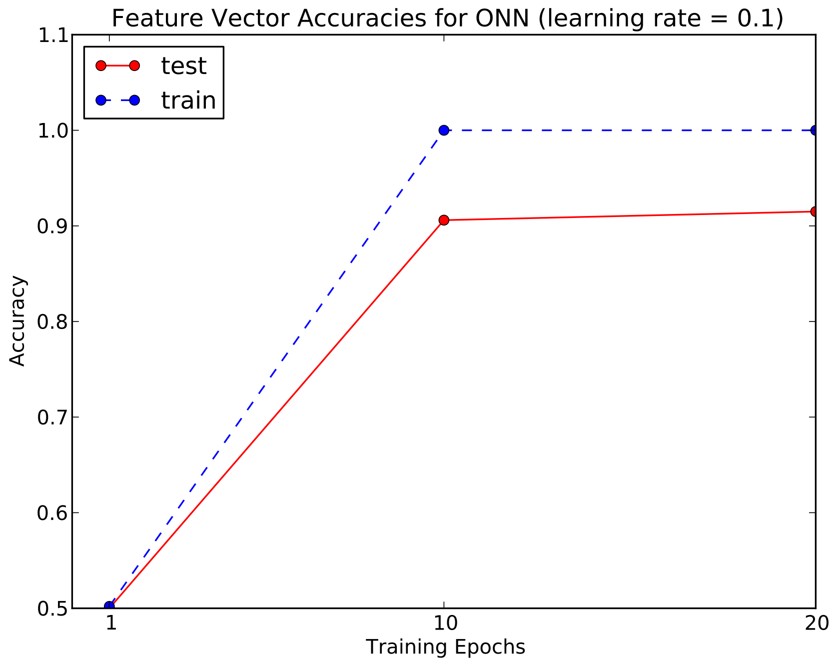
weight	update	final weight
w_{10}	$0 + 0.0127$	0.0127
w_{13}	$-2 + 0.00342$	-1.997
w_{14}	$1 + 0.00635$	1.0006
w_{20}	$-0.5 - 0.0012$	-0.5012
w_{23}	$-2 - 0.00032$	-2.0003
w_{24}	$-2 - 0.0006$	-2.0006
w_{30}	$-0.5 - 0.00452$	-0.505
w_{3x_1}	$1 + 0$	1
w_{3x_2}	$-0.5 - 0.00452$	-0.505
w_{40}	$-1 + 0.00378$	-0.996
w_{4x_1}	$-0.5 - 0$	-0.5
w_{4x_2}	$1 + 0.00378$	1.0038

2. An example neural network with hidden units can represent the given equation $y = (x_1 \wedge x_2) \vee (x_3 \wedge x_4)$ is shown below.

3. The three graphs for showing the training- and test-set accuracy are shown below. The training set accuracy is aggregated across each fold (e.g. I count the number of correct and incorrect guesses I get).







4. A sample ROC curve is shown below. There are only 4 false positives, so the graph isn't very gradual when it crosses the FPR spectrum!