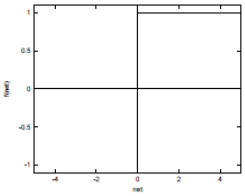
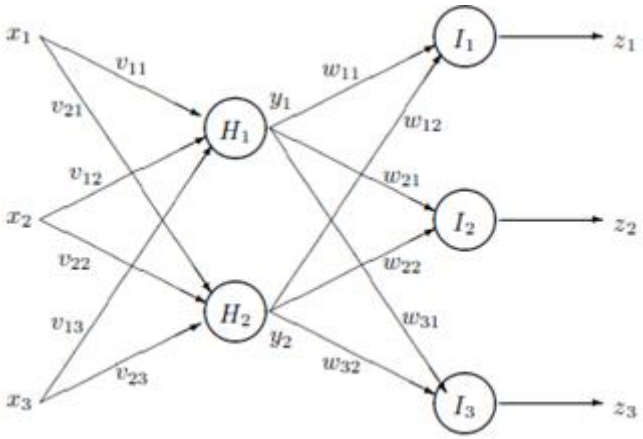
Class Exercises. Set 5

1. An artificial neural network may be trained on one data set and tested on a second data set. The system designer can then experiment with different numbers of hidden layers, different numbers of hidden units, etc. For real world applications, it is therefore important to use a third data set to evaluate the final performance of the system. Why?
A. The error on the third data set provides a better (unbiased) estimate of the true generalization error.
B. The error on the third data set is used to choose between lots of different possible systems.
C. It's not important: testing on the second data set indicates the generalization performance of the system.

2.
A perceptron with a unipolar step function has two inputs with weights w_1 = 0.5 and w_2 = -0.2, and a bias b = -0.3 (b can therefore be considered as a weight for an extra input which is always set to +1). For a given training example x = $[0; 1]^t$, the desired output is 0 (zero). Does the perceptron give the correct answer (that is, is the actual output the same as the desired output)?

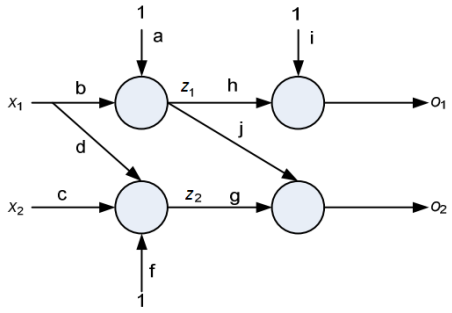

Unipolar step function

3. What is the usual sequence of events for training the network using the backpropagation algorithm?



a. (1) calculate $z_k = f(I_k)$, (2) update w_{kj} , (3) calculate $y_j = f(H_j)$, (4) update v_{ji}.
b. (1) calculate $y_j = f(H_j)$, (2) update v_{ji}, (3) calculate $z_k = f(I_k)$, (4) update w_{kj} .
c. (1) calculate $y_j = f(H_j)$, (2) calculate $z_k = f(I_k)$, (3) update v_{ji}, (4) update w_{kj} .
d. (1) calculate $y_j = f(H_j)$, (2) calculate $z_k = f(I_k)$, (3) update w_{kj} , (4) update v_{ji}.

4. Consider a multilayer perceptron with the structure indicated in the figure. In this figure, roman letters indicate weights and italic letters indicate network variables.



All weights have an initial value of 0.5. The output units are linear and the units in the first layer have, as activation function, the hyperbolic tangent.

The training set is

X_1	X_2	Y_1	Y_2
-1	1	1	1
1	-1	2	-2

a) Compute the output of all the 4 neurons.
b) Find, through the backpropagation method in **batch mode**, the value of weight **b** after one iteration with no momentum term and with a learning rate $\eta = 0.2$. The patterns are presented in the order given in the table, and the cost function is the total squared error.