4.1 What are the values of weights wo, wl, and w2 for the perceptron whose decision surface is illustrated in Figure 4.3? Assume the surface crosses the xl axis at -1, and the x2 axis at 2.

Ans. The function of the decision surface is: 2+2x1-x2 = 0, so w0 =2, w1 = 2, w2 = -1.

4.2. Design a two-input perceptron that implements the boolean function A $\wedge\neg$ B. Design a two-layer network of perceptrons that implements A XOR B.

Ans. We assume 1 for true, -1 for false.

(1) A $\wedge\neg$ B: w0 = -0.8. w1 = 0.5, w2 = -0.5.

x1(A)	x2(B)	w0+w1x1+w2x2	output
-1	-1	-0.8	-1
-1	1	-1.8	-1
1	-1	0.2	1
1	1	-0.8	-1

(2) A XOR B = (A $\wedge\neg$ B) $\vee$ ($\neg$A $\wedge$B)

The weights are:

Hidden unit 1: w0 = -0.8, w1 = 0.5, w2 = -0.5

Hidden unit 2: w0 = -0.8, w1 = -0.5, w2 = 0.5

Output unit: w0 = 0.3, w1 = 0.5, w2 = 0.5

x1(A)	x2(B)	Hidden unit 1 value	Hidden unit 2 value	Output value
-1	-1	-1	-1	-1
-1	1	-1	1	1
1	-1	1	-1	1
1	1	-1	-1	-1

4.3. Consider two perceptrons defined by the threshold expression $w_0 + w_1x_1 + w_2x_2 > 0$. Perceptron **A** has weight values: w0 = 1, w1 =2, w2 = 1. and perceptron B has the weight values: w0 = 0, w1 = 2, w2 = 1. True or false? Perceptron *A* is *more-general~than* perceptron B. *(more-general~than* is defined in Chapter 2.)

Ans. True.

For each input instance x=(x1, x2), if x is satisfied by B, which means 2x1+x2>0, then we have 2x1+x2 +1>0. Hence, x is also satisfied by the A.

4.5. Derive a gradient descent training rule for a single unit with output o, where

$$o = w_0 + w_1x_1 + w_1x_1^2 + \ldots + w_nx_n + w_nx_n^2$$

Ans.

The gradient descent is: $\nabla E(\vec{w}) = [\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w1}, \ldots, \frac{\partial E}{\partial wn}]$

$$\frac{\partial E}{\partial wi} = \frac{\partial}{\partial wi}\frac{1}{2}\sum_{d\in D}(t_d - o_d)^2$$
$$= \sum_{d\in D}(t_d - o_d)\frac{\partial}{\partial wi}(t_d - o_d)$$
$$= \sum_{d\in D}(t_d - o_d)(-x_{id} - x_{id}^2)$$

The training rule for gradient descent is: $w_i = w_i + \Delta w_i$, where

$$\Delta w_i = -\eta\frac{\partial E}{\partial w_i} = \eta\sum_{d\in D}(t_d - o_d)(x_{id} + x_{id}^2)$$