CNS 2013

Solutions to Bert Kappen Problem Set 4
Biophysics lecture notes - Chapter 6 - Exercise 2 & 3

Marije ter Wal

2. Equal convergence for 2D gradient descent

The gradients in both directions are:

$$\frac{\partial E}{\partial x} = 2a_1 x \tag{1}$$
$$\frac{\partial E}{\partial y} = 2a_2 y \tag{2}$$
$$\tag{3}$$

After one 'step' of the algorithm the new value for x is:

$$\begin{aligned} x_{i+1} &= x_i - \eta\frac{\partial E}{\partial x} && (4)\\ \Rightarrow \Delta x_{i+1} &= -\eta 2a_1 x_i && (5)\\ &= -\eta 2a_1(x_{i-1} - \eta 2a_1 x_{i-1}) && (6)\\ &= -\eta 2a_1 x_{i-1}(1-\eta 2a_1) && (7)\\ &= \Delta x_i(1-\eta 2a_1) && (8)\end{aligned}$$

Similarly,

$$\Delta y_{i+1} = \Delta y_i(1-\eta 2a_2) \tag{9}$$

We require the convergence, so the absolute change in the change of x and y, to be identical, hence:

$$|1-\eta 2a_1| = |1-\eta 2a_2| \tag{10}$$

As Figure shows, for the solution holds:

$$\begin{aligned} -(1-\eta 2a_1) &= (1-\eta 2a_2) && (11)\\ \Rightarrow \eta &= \frac{1}{a_1+a_2} && (12)\end{aligned}$$

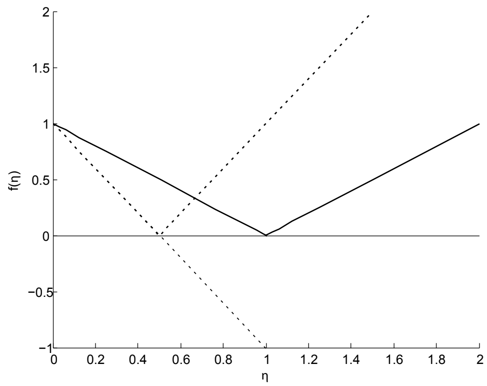

Fig. 1: Equation 10 plotted.

3. Perceptron AND operator

a. Weights and threshold

Due to symmetry we can assume: $w_1 = w_2 = w$. The input-output pairs for learning are:

$$\begin{aligned} \xi^\mu &= (\xi_0, \xi_1, \xi_2) && (13)\\ \xi^1 &= (-1,-1,-1); \zeta^1 = -1 && (14)\\ \xi^2 &= (-1,-1,+1); \zeta^2 = -1 && (15)\\ \xi^3 &= (-1,+1,-1); \zeta^3 = -1 && (16)\\ \xi^4 &= (-1,+1,+1); \zeta^4 = +1 && (17)\\ & && (18) \end{aligned}$$

The quadratic cost function is than:

$$\begin{aligned} E(w) &= \frac{1}{2}\sum_\mu \left(\zeta^\mu - \sum_j w_j \xi_j^\mu\right)^2 && (19)\\ &= \frac{1}{2}\left((-1+w_0+2w)^2 + 2(-1+w_0)^2 + (1+w_0-2w_1)^2\right) && (20)\\ &= w_0^2 + (1-2w)^2 + (-1+w_0)^2 && (21) \end{aligned}$$

Computing the derivatives of this expression with respect to w_0 and w and setting them to 0 yields the optimal values for both the weights and the threshold:

$$\frac{\partial E}{\partial w_0} = 2w_0 + 2(-1 + w_0) \tag{22}$$
$$= -2 + 4w_0 \tag{23}$$
$$\Rightarrow w_0 = \frac{1}{2} \tag{24}$$

$$\frac{\partial E}{\partial w} = -4(1 - 2w) \tag{25}$$
$$= -4 + 8w \tag{26}$$
$$\Rightarrow w = \frac{1}{2} \tag{27}$$

Filling in these values in the cost function:

$$E(w) = w_0^2 + (1 - 2w)^2 + (-1 + w_0)^2 \tag{28}$$
$$= \frac{1}{2}^2 + (1 - 2\frac{1}{2})^2 + (-1 + \frac{1}{2})^2 \tag{29}$$
$$= \frac{1}{2} \tag{30}$$

b. Linear dependence

For linearly independent inputs, the patterns can be separated perfectly, the second term in the sum of the cost function will be equal to the first one and hence $E = 0$. Any $E > 0$ indicates a imperfect separation and hence a linear dependence.