1. Find explicitly the coefficients for the quadratic threshold gate \(Q(x, y) = \omega_0 + \omega_1 x + \omega_2 y + \omega_{12} xy + \omega_{11} x^2 + \omega_{22} y^2\) which solves the XOR function with one neuron. Hint: solve it first when the domain of input signal is \((-1, 1)^2\) (instead of \((0, 1)^2\)) and hence \(XOR(-1, -1) = 0\) and \(XOR(1, -1) = XOR(-1, 1) = 1\). Try to consider all 6 cases discussed in the lectures. Show that 14 out of 16 Boolean switching functions (for \(n = 2\)) are linearly separable. Is it possible to show the counterexample in case of \(n = 2\) of non-quadratically separable function \(f\): a) for \(f\) being Boolean b) for \(f\) being an arbitrary function \(R^2 \to R\).

2. Check whether the entropy energy function \(J : \Omega \to R \geq 0\) defined over \(\Omega \subset R^{n+1}\)

\[
J(\omega) = 1/2 \sum_{i=1}^{m} \left[ (1 + d^i) \ln \frac{1 + d^i}{1 + y^i} + (1 - d^i) \ln \frac{1 - d^i}{1 - y^i} \right]
\]

for \(y^i \in (-1, 1)\) and \(d^i \in (-1, 1)\) satisfies the condition \(J(\omega) \geq 0\). Here \(d^i\) (for \(1 \leq i \leq m\)) denotes the desired net value for \(x^i = (x_1^i, x_2^i, ..., x_n^i, x_{n+1}^i = 1)\) and \(y^i = \phi(\sum_{j=1}^{n+1} \omega_j x_j^i)\) denotes the net output while \(x^i\) is input and \(\phi\) being some activation function. Note that \(J\) vanishes for \(y^i = d^i\) (for each \(1 \leq i \leq m\)). The function \(J\) can be used as and alternative to the Least Square Error energy function \(E\) commonly applied in ANN. The main advantage of \(J\) over \(E\) (if applied) is to eliminate the so-called flat spots (when \(\phi'\) is small) which slow down the convergence of the gradient descent algorithm - see lecture notes). More specifically if the activation function is tansgosoidal, namely \(\phi(x) = g_\beta(x)\) (see Tutorial 1) verify that the gradient \(\nabla J\) does not contain the term \(g'_\beta(x)\), which for \(|x| > 3\), is close to zero and makes the gradient descent algorithm very inefficient. It is sufficient if you consider only the case when \(m = 1\).

3. Apply a simple correction rule (for \(\rho = 1/2\) and thus \((w^{k+1} = w^k \pm x^k\) if \(x^k\) is misclassified by current \(w^k\), with + or − taken accordingly) in the perceptron algorithm with \(w^0 = (0, 0, 0, 0)\) for the Boolean function \(f(0, 1, 1) = -1, f(0, 0, 0) = 1, f(1, 0, 0) = 1, f(1, 0, 1) = 1, f(0, 0, 1) = -1, f(1, 1, 0) = 1, f(1, 1, 1) = -1\) and \(f(0, 1, 0) = -1\). Check that 29 pattern presentations are needed and 13 effective weight adjustments eventuate.

4. Given a correction factor chosen as \(\rho = 1/4\), \(w^0 = (1, -1, 0)\) (an initial starter) and the training set of vectors \(X^1 = (1, 2), X^2 = (1, 3), X^3 = (1, 0)\) and \(X^4 = (1, -1)\) for which only the the last belongs to Class A pattern (+1) and the first three to the Class B pattern (-1) check that a) the training set \(X\) is linearly separable b) show consecutive steps of the perceptron algorithm c) show the current update for the decision boundary once the \(w^k\) is being changed. Apply here a simple correction rule.

Assoc. Prof. Ryszard Kozera