Problem Set 3

Submission instructions and due date:

The due date for this homework assignment is as listed on Blackboard as part of the "Problem Set ..." file name on Blackboard.

The completed assignment must be submitted in class (on paper and stapled according to the instructions below) at the beginning of the class session on the above due date, unless different submission instructions are announced in class or on Blackboard.

What to submit (print and staple in this order):

1) hw3scores.txt (complete and print out).
2) solutions to written problems.

For item 2 of the list above (solutions to written problems) – solutions to written problems can be handwritten; however, they should be written out clearly. It is required that you show all relevant steps leading to the solutions.

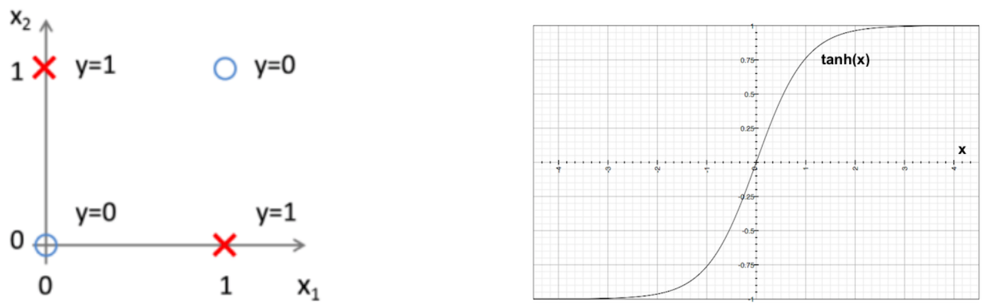
Grading:

This problem set is self-graded. It is strongly recommended that you solve and complete all of its tasks, since – in addition to contributing towards the homework component of the final course-grade – they should help you prepare for quizzes/exams. You must also submit your self-assigned grades in "hw3scores.txt" (see also instructions within that file). To verify self-grading, for each homework assignment we will randomly choose a subset of submissions for grade checking and re-grade them if needed.

Reminders: Copying solutions from any source – e.g., online, solutions manuals, assignment solutions from prior semesters, or solutions from any other sources – is prohibited.

1 Written Problem 1 [25 points].

Design a neural network to solve the XOR problem, i.e. the network should output 1 if only one of the two binary input variables is 1, and 0 otherwise (see in Figure 1a). Use the hyperbolic tangent, or *tanh*, activation function in all nodes (Figure 1b), which ranges in [-1, +1].



(a) (b)

Figure 1

Note that (A XOR B) can be expressed as (A OR B) AND NOT (A AND B), as illustrated below (see in Figure 2):

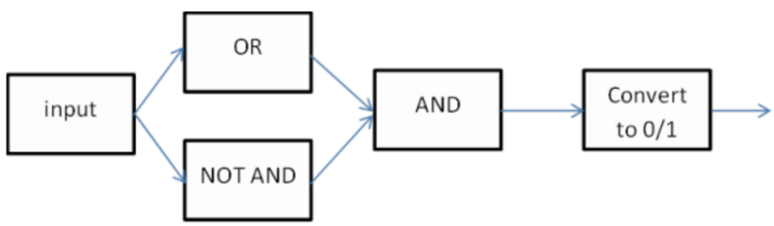


Figure 2

In the diagrams shown in Figure 3, we filled in most of the tanh units' parameters. Fill in the remaining parameters, keeping in mind that tanh outputs +1/-1, not 0/1. Note that we need to appropriately change the second layer (the AND node) to take +1/-1 as inputs. Also, we must add an extra last layer to convert the final output from +1/-1 to 0/1.
Hint: Assume that tanh outputs -1 for any input $x \leq -2$, *+1 for any input* $x \geq +2$, *and 0 for* $x = 0$.

2 Written Problem 2 [25 points].

Show that maximizing the likelihood function under the conditional distribution

$$p(t|x, w) = N(t|y(x, w), \beta^{-1}I)$$

for a multi-output neural network is equivalent to minimizing the sum-of-squares error function

$$E(w) = \frac{1}{2}\sum_{n=1}^{N} \|y(x_n, w) - t_n\|^2.$$

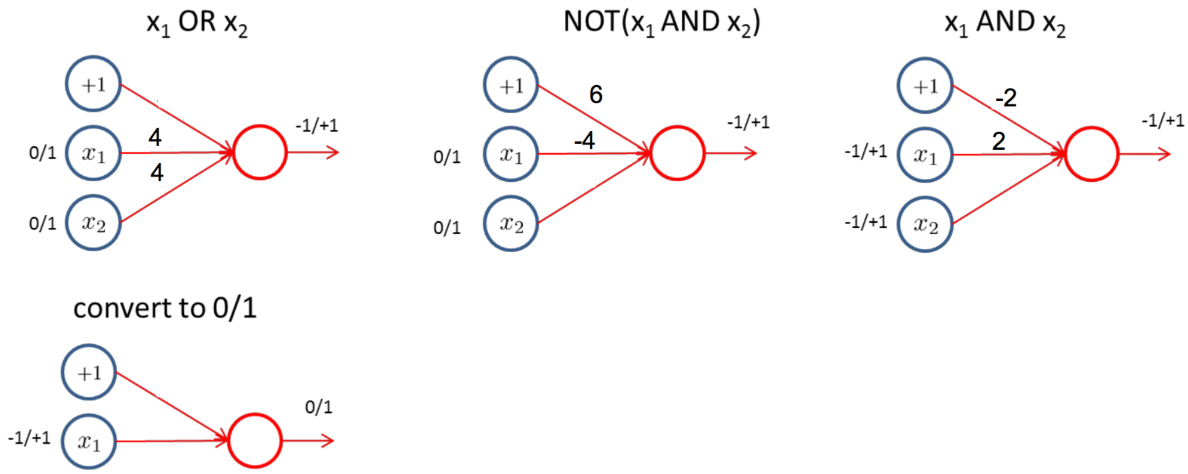


Figure 3

3 Written Problem 3 [25 points].

Consider a multiclass problem in which each input is assigned to one of K mutually exclusive classes, the binary target variables $t_k \in \{0, 1\}$ having a 1-of-K coding scheme indicating the class, and consider a multiclass neural network model in which the network outputs have the interpretation $y_k(x, w) = p(t_k = 1|x)$. Show that maximizing likelihood for this network model leads (is equivalent) to the minimization of the following cross-entropy error function:

$$E(w) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \ln y_k(x_n, w)$$

4 Written Problem 4 [25 points].

The (cross-entropy) error function for binary classification problems for a network having a logistic-sigmoid output activation function, so that $0 \leq y(x, w) \leq 1$, and data having target values $t \in \{0, 1\}$, can be shown to be of the form:

$$E(w) = -\sum_{n=1}^{N} \{t_n \ln y_n + (1 - t_n) \ln(1 - y_n)\}$$

Derive the corresponding error function if we consider a network having an output $-1 \leq y(x, w) \leq 1$ and target values $t = 1$ for class C_1 and $t = -1$ for class C_2. What would be the appropriate choice of output unit activation function?