CMPS 142 Fourth Homework, Winter 2016

2 Problems, 11 pts, due start of class Wednesday Feb. 24

This homework is to be done in groups of 2 or 3. Each group members should completely understand the group's solutions and *must* acknowledge all sources of inspiration, techniques, and/or helpful ideas (web, people, books, etc.) other than the instructor, TA, and class text. Each group should submit a single set of solutions containing the names and e-mail addresses of all group members. Although there are no points for "neatness", the TA may deduct points for illegible or poorly organized solutions.

1. (5 pts) Consider learning a decision tree with boolean features where the impurity criterion is the error (misclassification) rate, so if a set of 10 examples had 3 positive and 7 negative example, the impurity would be 3 / 10.

   First (1 pt), show that the average impurity of a split is the same as the fraction of mistakes made if predictions were made with the majority class at each node on the next level. (recall from the lecture slides that the average impurity of a split of $n$ examples into $n_1$ and $n_2$ example subsets is

$$\frac{n_1(\text{ impurity of the } n_1 \text{ subset}) + n_2(\text{ impurity of the } n_2 \text{ subset})}{n}$$

   Therefore the attribute resulting in the fewest errors at the next level (when each subtree's examples are predicted with the majority label in the subtree) will be selected as the test.

   Second (4 pts), Construct a simple training set with binary features and labels where:

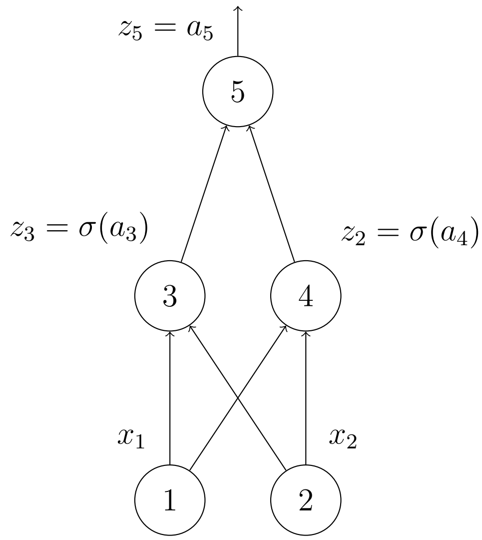
   (a) There is small decision tree that correctly classifies the training data,

   (b) The greedy decision tree construction algorithm finds a larger tree (i.e. more nodes), and

   (c) There are no ties in the construction process (i.e. the decision tree algorithm finds a single best attribute to test at each node).

   List your training set and draw the small decision tree that correctly classifies it. Illustrate how the greedy decision tree algorithm runs on your data, and show the resulting tree. You may duplicate examples in your training set if you find it helpful. For all four points, construct a training set where the tree found by the decision tree algorithm has more than twice as many nodes as the smallest tree correctly classifying the data.

2. (6 pts) Consider the following artificial neural network.




Recall that each node as an $a$-value that is a weighted sum of the nodes inputs, and produces a $z$ (output value).

For the first part, let the $\sigma()$ function be the logistic sigmoid, $\sigma(a) = 1/(1 + e^{-a})$. Assume that the the nodes do *not* have bias terms, and the initial weights are all 0's, the error on the output is the squared error, $\frac{1}{2}(z_5 - y)^2$, so $z_5 = a_5$ (there is no sigma-function at the output node). Under these assumptions, perform one step of backpropagation (by hand) with step size $\eta = 0.1$ on the training example $x_1 = 1$, $x_2 = 2$, $y = 1$. Show the $a_i$, $z_i$ and $\delta_i$ values for each non-input node, and the new weights after the backprop update. See the backpropagation handout for the procedure to use.

For the second part, assume that the units are rectified linear units, so $\sigma(a) = \max(0, a)$. Assume that there are no bias terms, and the initial weights are (recall $w_{i.j}$ is the weight at node $i$ on incoming value $z_j$).

$$
\begin{aligned}
w_{5,3} &= 1 \\
w_{5,4} &= 1 \\
w_{4,2} &= -1 \\
w_{4,1} &= 1.5 \\
w_{3,2} &= 1 \\
w_{3,1} &= -0.5
\end{aligned}
$$

Again using the squared error, $\frac{1}{2}(z_5 - y)^2$, perform one iteration of backpropagation on example $x_1 = 1$, $x_2 = 2$, $y = 1$ with step size $\eta = 0.1$. Show the $a_i$, $z_i$ and $\delta_i = \partial \text{error}/\partial a_i$ values for each non-input node, and the new weights after the backprop update.