CSE 250B Quiz 9

Tuesday March 11, 2013

Instructions. Do this quiz in partnership with exactly one other student. Write both your names at the top of this page. **Circle one name that we will call out when we return the quiz. Choose a first name or last name that is likely to be unique in the class.**

Discuss the answer to the question with each other, and then write your joint answer below the question. Use the back of the page if necessary. It is fine if you overhear what other students say, because you still need to decide if they are right or wrong. You have seven minutes. The maximum score is three points.

Question. Consider a neural network that is a binary tree with three input nodes, one hidden node, and one output node. The value of every node is a scalar. Assume that

- the input node values are $x_1 = x_2 = x_3 = 1$
- both non-input nodes have the **same** shared weight matrix $W = [w, v] \in \mathbb{R}^2$
- the hidden node has transfer function h while the output node is linear
- the target output is $t = 0$ and the loss function $J(z, t) = (z - t)^2$ is quadratic where z is the value of the output node.

Obtain an expression for $\partial J / \partial w$.

Answer. The value of the hidden node is $h(w+v)$. The value of the output node is $z = w + vh(w+v)$ if node 1 feeds directly into the output, or $z = wh(w+v) + v$ if node 3 does. We'll assume the former. The partial derivative is

$$\frac{\partial}{\partial w} J = \frac{\partial}{\partial w} z^2 = 2z \frac{\partial}{\partial w} z = 2z[1 + vh'(w+v)].$$

Additional note. It is tractable to calculate partial derivatives symbolically only for simple networks like the one above. In contrast, backpropagation allows derivatives to be computed efficiently for all networks.

In general, for an edge from node i to node j, $\partial J/\partial w_{ij} = \delta_j z_i$ with $\delta_j = \partial J/\partial a_j$ where a_j is the activation coming into node j. Here, for the output node z, $\delta_z = \partial J/\partial a_z = \partial z^2/\partial z = 2z$ and $\partial J/\partial w = 2zx_1 = 2z$. For the hidden node h, $\delta_h = h'(a_h)\delta_z v$ by backpropagation, so $\partial J/\partial w = h'(a_h)2zvx_2$. When a weight w is used twice, the actual partial derivative with respect to it is the sum of the partial derivatives from each place where the weight is used. Here, this sum is $2z + h'(a_h)2zv = 2z[1 + h'(w + v)v]$.

This result is the same as computed directly above. When two different ways of computing the same quantity lead to the same answer, that provides confidence (but not certainty) that both ways are correct.