Issued: February 25, 2010 IE789F/OR791S HW 5 Due: March 9, 2010

Exercises

1. You are using a simple three-layer neural network to predict (a) the total sales (in millions of dollars) (b) inventory level (in days of finished products) and (c) yield of products (in percentage) of a large manufacturing company. The neurons of the output layer are using the conventional $g(u)$ activation function. How can you ask the neural network to predict meaningful numbers for your three outputs at the same time? What are potential issues in training the network? How will you take care of these issues?

2. Given
$$h(u) = \frac{1 - e^{-u}}{1 + e^{-u}}$$
and
$$\tanh(\mathrm{u}) = \frac{\mathrm{e^u} - \mathrm{e^{-u}}}{\mathrm{e^u} + \mathrm{e^{-u}}}$$
   (a) Find an interval around zero and prove the slope of $\tanh(u)$ is deeper than that of $h(u)$ in the interval.
   (b) How to make the slope even deeper?
   (c) Why do you want to make the slope deeper?

3. Replace g$(\cdot)$ by h$(\cdot)$ and, as in our classroom, derive $\frac{\partial \bar{E}}{\partial b_{jk}}$ $\frac{\partial \bar{E}}{\partial a_{ij}}$ and $\frac{\partial \bar{E}}{\partial c_{ik}}$, where c_{ik} is the weight on the link directly connecting input node i to output node k.

4. Replace g$(\cdot)$ by $\tanh(\cdot)$ and derive $\frac{\partial \bar{E}}{\partial b_{jk}}$ and $\frac{\partial \bar{E}}{\partial a_{ij}}$ and $\frac{\partial \bar{E}}{\partial c_{ik}}$.

5. Consider the two-dimensional function
$$f(\mathbf{x}) = (\mathbf{x_1^2} + \mathbf{x_2^2} - \mathbf{1})^\mathbf{2}$$
over the domain $\Omega = \{(x_1, x_2)^T \mid -1 \leq x_1 \leq 1,\ -1 \leq x_2 \leq 1\}$. Your assignment is to create a simple three-layer neural network to approximate this function using randomly selected 100 data points of $\{(x, f(x)\}$. Write a program (matlab is ok) to show how good your neural network can approximate the target function. You can randomly select another 100 points in Ω to draw your output pictures. Remember to report the process of your design/data/ training, validate your results using a table of (input vector, target value, approximate value), and discuss your experience in solving this problem (such as using $h(\cdot)$ vs. $g(\cdot)$ etc.)