Neural Network Assignment 2: (2 parts), Professor Q.J. Zhang, Carleton University

Part 1: Perform Training Numerically

For a MLP network with 1 input neuron, 2 hidden neurons with Sigmoid functions, and 1 output neuron with linear function as shown in Figure A1, train the neural network using batch-mode backpropagation method with fixed learning rate $\eta = 0.1$ and no momentum. Specifically, using the initial values of weights in Figure A1, calculate the new weight values for 1 epoch. In the calculation, assume no scaling to x or y, and the training error is defined as

$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{P}\left(y(x_i, \boldsymbol{w}) - d_i\right)^2$$

where (x_i, d_i) represents the i^{th} sample of training data, i=1, 2, …, P, and P is the total number of training samples. For this example, P=2 and the training data is shown in Table A1.

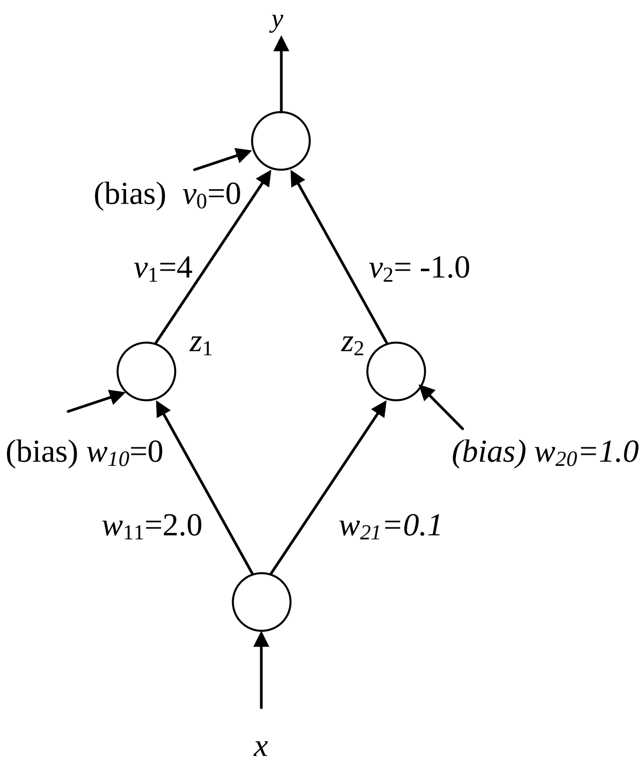

Figure A1

Table A1
Training Data

i	x_i	d_i
1	0	2
2	1	4

Show the values and corresponding computations for E, the derivatives $dE/d\boldsymbol{w}$, and the $\boldsymbol{w}$ in the first epoch. (Make sure to include the updated values of $\boldsymbol{w}$ at end of the epoch).

Part 2: Perform Training by Graphical Illustration

The purpose of this exercise is to use graphical illustration to demonstrate neural network training algorithms. It is suggested that a ruler (or a graph paper with grids) is used to draw the graphs.

Let E($\boldsymbol{w}$) represent the training error as a function of neural network internal weights**,** where $\boldsymbol{w}$ is a vector of neural network internal weights. For convenience of graphical explanation, we assume that we consider only two variables for neural network training, and the function E($\boldsymbol{w}$) is simplified using 2nd order Taylor series.

More specifically, let E($\boldsymbol{w}$) for batch mode training be described as

$$E(\boldsymbol{w}) = (w_1)^2 - 2w_1 + 1 + 2((w_2)^2 - 4w_2 + 4) \qquad (1)$$

where $\boldsymbol{w}$ is a vector of two variables

$$\boldsymbol{w} = [w_1 \quad w_2]^T \qquad (2)$$

and superscript T denotes the transpose of the vector.

Suppose the initial value of $\boldsymbol{w}$ for neural network training is:

$$\boldsymbol{w} = [0.5 \quad 1.0]^T \qquad (3)$$

Use graphical illustration to carry out two epochs of neural network training with batch-mode backpropogation (part 2a) and conjugate gradient (part 2b) methods.

Part 2(a): You are required to:

1. Draw the contour plot of the E($\boldsymbol{w}$) in the 2-dimensional $\boldsymbol{w}$ space.
2. On the contour plot, indicate the initial point of $\boldsymbol{w}$
3. On the contour plot, show the gradient direction ∇E (where $\nabla E = \frac{\partial E}{\partial \boldsymbol{w}}$)
4. On the contour plot, show the direction $\boldsymbol{h}$ for the batch mode backpropagation method assuming the momentum factor is zero.
5. On the contour plot, show the new location of $\boldsymbol{w}$ after one epoch of training is finished, assuming we have used line minimization to determine the optimal step size η.
6. Repeat 3-5 above for one more epoch. Indicate the new location of $\boldsymbol{w}$ at end of the 2nd epoch.

Part 2(b): Suppose that we use Conjugate Gradient method to do the training where E($\boldsymbol{w}$), $\boldsymbol{w}$ and initial values of $\boldsymbol{w}$ are defined as in (1)-(3).

1. How many epochs are needed to find the optimal solution of ***w*** so that the training error is reduced to zero ?
2. Draw the contour plot of E(***w***) in the 2-dimensional ***w*** space, indicate the initial point of ***w*,** indicate the update direction ***h*,** indicate the new location of ***w*** at end of the first epoch.
3. Show the negative gradient direction (i.e., $-\nabla E$) in the 2^{nd} epoch.
4. Show the conjugate gradient direction on the contour plot in the 2^{nd} epoch.
5. Indicate the new location of ***w*** at end of 2^{nd} epoch.