CS 4804 Homework 6
Solution Sketches

---

1. **(20 points)** The error for a single example can be given by

$$\begin{aligned} E &= \frac{1}{2}Err^2 \\ &= \frac{1}{2}(y - \frac{1}{2b}(x_1w_1 + x_2w_2 + x_3w_3 + x_4w_4 + b))^2 \\ &\quad \text{if} -b \leq x_1w_1 + x_2w_2 + x_3w_3 + x_4w_4 \leq b \end{aligned}$$

   Differentiating with respect to each weight gives:

$$\begin{aligned} \frac{\partial E}{\partial w_i} &= -Err \times \frac{1}{2b}\, x_i \\ \frac{\partial E}{\partial b} &= Err \times (\frac{x_1w_1}{2b^2} + \frac{x_2w_2}{2b^2} + \frac{x_3w_3}{2b^2} + \frac{x_4w_4}{2b^2}) \end{aligned}$$

   Which gives learning rules of the form:

$$w_i = \begin{cases} w_i + \alpha \times Err \times x_i & \text{if } x_1w_1 + x_2w_2 + x_3w_3 + x_4w_4 < -b \\ w_i + \alpha \times Err \times x_i & \text{if } x_1w_1 + x_2w_2 + x_3w_3 + x_4w_4 > b \\ w_i + \alpha \times Err \times (\frac{1}{2b}\, x_i) & \text{otherwise} \end{cases}$$

   and

$$b = \begin{cases} b + \alpha \times Err & \text{if } x_1w_1 + x_2w_2 + x_3w_3 + x_4w_4 < -b \\ b - \alpha \times Err & \text{if } x_1w_1 + x_2w_2 + x_3w_3 + x_4w_4 > b \\ b - \alpha \times Err \times (\frac{x_1w_1}{2b^2} + \frac{x_2w_2}{2b^2} + \frac{x_3w_3}{2b^2} + \frac{x_4w_4}{2b^2}) & \text{otherwise} \end{cases}$$

   This will clearly have multiple local maximum and minimum as a simple plot of the error surface will show.

2. **(20 points)** The only modification that must be made to the standard back propagation algorithm is that the output node must use not only the hidden node outputs, but also all the inputs. One possible set of weights is as follows:

   $y = g(0.438 \times g(0.062x_1 + 0.024x_2 + 0.157x_3 + 0.143x_4) - 0.352x_1 - 0.824x_2 + 0.737x_3 + 0.748x_4)$

   Other equations are also possible, depending on your initial weight choices.

3. **(60 points)** This exercise requires a careful implementation of the backpropagation algorithm. One possible formulation of the problem is to use a local encoding of Iris-setosa = 0.1, Iris-versicolor = 0.5, and Iris-virginica = 0.9 on one output node with three hidden nodes. Keep in mind that one of the classes is linearly separable but not the other two, so some number of hidden layer nodes is required.