1 Optimization

We would like to classify some data. We have $N$ samples, where each sample consists of a feature vector $\mathbf{x} = \{x_1, \cdots, x_k\}$ and a label $y = \{0, 1\}$.

We introduce a new type of classifier called logistic regression, which produces predictions as follows:

$$P(Y = 1 | X) = h(\mathbf{x}) = s \left( \sum_i w_i x_i \right) = \frac{1}{1 + \exp(-\sum_i w_i x_i)}$$

$$s(\gamma) = \frac{1}{1 + \exp(-\gamma)}$$

where $s(\gamma)$ is the logistic function, $\exp x = e^x$, and $\mathbf{w} = \{w_1, \cdots, w_k\}$ are the learned weights.

Let’s find the weights $w_j$ for logistic regression using stochastic gradient descent. We would like to minimize the following loss function for each sample:

$$L = -[y \ln h(\mathbf{x}) + (1 - y) \ln(1 - h(\mathbf{x}))]$$

(a) Find $dL/dw_j$. Hint: $s'(\gamma) = s(\gamma)(1 - s(\gamma))$.

(b) Write the stochastic gradient descent update for $w_j$. Our step size is $\eta$.  


Q2. Backpropagation

(a) Perform forward propagation on the neural network below for \( x = 1 \) by filling in the values in the table. Note that (i), \ldots, (vii) are outputs after performing the appropriate operation as indicated in the node.

\[
\begin{array}{ccccccc}
(i) & (ii) & (iii) & (iv) & (v) & (vi) & (vii) \\
\end{array}
\]

(b) Below is a neural network with weights \( a, b, c, d, e, f \). The inputs are \( x_1 \) and \( x_2 \).

The first hidden layer computes \( r_1 = \max(c \cdot x_1 + e \cdot x_2, 0) \) and \( r_2 = \max(d \cdot x_1 + f \cdot x_2, 0) \).

The second hidden layer computes \( s_1 = \frac{1}{1 + \exp(-a \cdot r_1)} \) and \( s_2 = \frac{1}{1 + \exp(-b \cdot r_2)} \).

The output layer computes \( y = s_1 + s_2 \). Note that the weights \( a, b, c, d, e, f \) are indicated along the edges of the neural network here.

Suppose the network has inputs \( x_1 = 1, x_2 = -1 \).

The weight values are \( a = 1, b = 1, c = 4, d = 1, e = 2, f = 2 \).

Forward propagation then computes \( r_1 = 2, r_2 = 0, s_1 = 0.9, s_2 = 0.5, y = 1.4 \). Note: some values are rounded.

Using the values computed from forward propagation, use backpropagation to numerically calculate the following partial derivatives. Write your answers as a single number (not an expression). You do not need a calculator. Use scratch paper if needed.

\[ g(z) = \frac{1}{1 + \exp(-z)} \text{, the derivative is } \frac{\partial g}{\partial z} = g(z)(1 - g(z)). \]

\[
\begin{array}{cccccccc}
\frac{\partial y}{\partial a} & \frac{\partial y}{\partial b} & \frac{\partial y}{\partial c} & \frac{\partial y}{\partial d} & \frac{\partial y}{\partial e} & \frac{\partial y}{\partial f} \\
\end{array}
\]