Predicting Admissions for Liver Disease

We consider a classification problem in which individuals are being assessed for possible liver disease treatment. The set of diagnostic quantities obtainable for each person are:

1. Age
2. Sex at birth
3. Total Bilirubin
4. Direct Bilirubin
5. Alkaline Phosphotase
6. Alamine Aminotransferase
7. Aspartate Aminotransferase
8. Total Protiens
9. Albumin
10. Albumin to Globulin Ratio

The first two quantities are obtained from the patient and the remaining eight quantities are obtained through standard blood sample analysis. A sample data set and description is obtainable at https://archive.ics.uci.edu/dataset/225/ilpd+indian+liver+patient+dataset. This set contains data for 583 people examined for liver disease. Each person is also associated with a class label, determined by an expert physician, identifying whether or not liver disease is indicated. Each quantity is measured in standard units with age in years (integer) and sex as a binary indicator 'F' or 'M'. The remaining quantities are non-negative real values.

The goal of this project is to construct and test one or more "feed-forward neural networks" (FFNN) that can be used to predict indicated liver disease in new potential patients (as a supplement to expert opinion). A FFNN is an example of a classifier function $y_k = f(x_k; w)$ where $x_k \in \mathbb{R}^p$ is a feature vector of diagnostic quantities associated with person k, $y_k \in \mathbb{R}$ is the classification value for person k, and $w \in \mathbb{R}^m$ is a parameter vector. That is, we assume that a good classifier is among the family of functions specified by a particular parameter choice. We wish to solve

$$\min_{w} \quad F(w) = \frac{1}{2} \sum_{k=1}^{n} \left(y_k - f(x_k; w) \right)^2 .$$

A FFNN can be represented as

$$f(x) = \sigma \left(W_d \sigma \left(\cdots \sigma \left(W_2 \sigma \left(W_1 x \right) \right) \cdots \right) \right),$$

where W_j is a $r_j \times c_j$ matrix (with $r_j = p$, $c_j = r_{j+1}$, and $c_d = 1$) and σ is the (elementwise) sigmoid function

$$\sigma(u) = \frac{1}{1 + \exp(-u)}.$$

The parameters of the function are contained in the weight matrices $W_1, \ldots, W_d$. That is, $m = \sum_{j=1}^{d} (r_j c_j)$. The gradient of f can be efficiently computed using the method of backpropagation. Let G_j be the matrix of gradient vector values associated elementwise with the weights in W_j. Then, we have the forward computation:

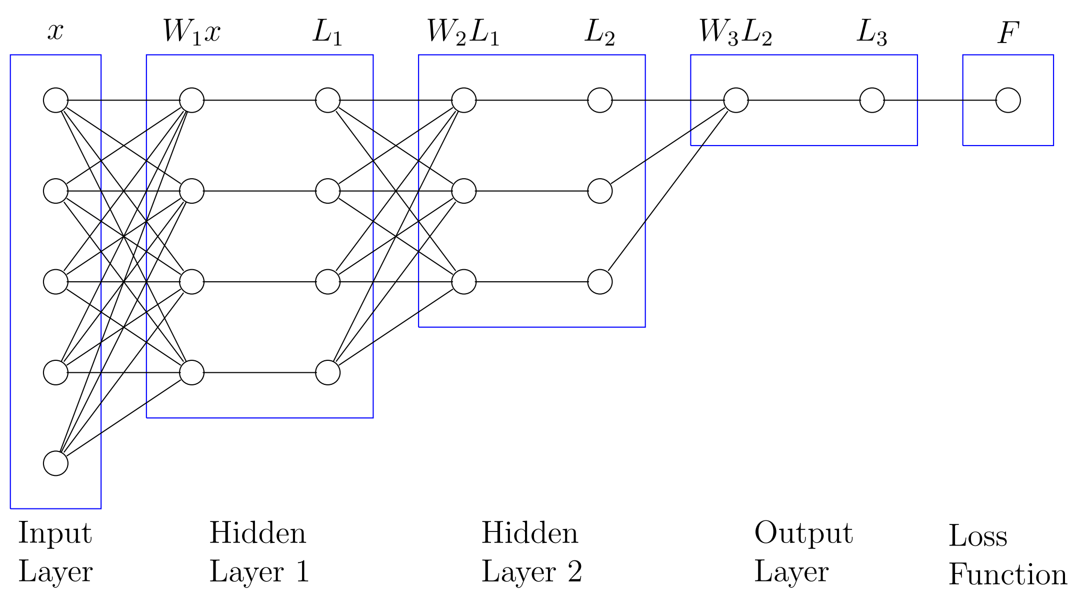
$$\begin{aligned} L_1 &= \sigma(W_1 x) \\ L_2 &= \sigma(W_2 L_1) \\ &\vdots \\ L_d &= \sigma(W_d L_{d-1}) \\ F &= \tfrac{1}{2} \left\| L_d - y \right\|^2 \end{aligned}$$

and the gradient computation (using the multivariate chain rule):

$$\begin{aligned} h_d &= (L_d - y) L_d (1 - L_d), \qquad G_d = h_d L_{d-1}^{\mathsf{T}} \\ h_{d-1} &= W_d^{\mathsf{T}} h_d L_{d-1}(1 - L_{d-1}), \qquad G_{d-1} = h_{d-1} L_{d-2}^{\mathsf{T}} \\ &\vdots \\ h_2 &= W_3^{\mathsf{T}} h_3 L_2 (1 - L_2), \qquad G_2 = h_2 L_1^{\mathsf{T}} \\ h_1 &= W_2^{\mathsf{T}} h_2 L_1 (1 - L_1), \qquad G_1 = h_1 x^{\mathsf{T}} \end{aligned}$$

In this derivation, we have used the fact that $\sigma'(u) = \sigma(u)(1 - \sigma(u))$. Constructing the problem in terms of matrices W_j and G_j provides convenient notation. However, it must be remembered that the parameter vector $w \in \mathbb{R}^m$ is constructed from the collective and ordered elements of $W_1, W_2, \ldots, W_d$. Similarly, the gradient vector $g \in \mathbb{R}^m$ is constructed from the identically ordered elements of $G_1, G_2, \ldots, G_d$. One interesting point is that if one employs the identity function instead of the sigmoid (or other nonlinear monotonic function) then the entire transformation is linear with both forward and backward computations collapsing into a single matrix product.

It can be helpful to visualize the FFNN as a computational directed acyclic graph. The forward computation proceeds left to right.



In this example, the network contains two hidden layers and uses three weight matrices. W_1 is a 4×5 matrix, W_2 is a 3×4 matrix, and W_3 is a 1×3 matrix. The computations within each layer are the elementwise sigmoid "activation" functions. The input x can be a single vector of (five, in this example) attributes, or it can be an array (five by m) of attributes of m individuals.

Complete the following tasks.

1. Construct a function which reads the liver disease data set, removes persons for which some data is missing, converts each person's sex to a binary value, and normalizes each diagnostic quantity to the maximum present in the data.

2. Construct a function that computes – for a general, user-specified FFNN – the loss function for a classification problem and the (backpropagation) gradient.

3. Choose a method for selecting training data and test data sets.

4. Solve the classification problem using a FFNN which includes all 10 input quantities and two hidden layers, each with 10 nodes. Test various optimization methods using the code you have developed.

5. Explore other choices of FFNN construction.