Description of dataset, response variable, and explanatory variables

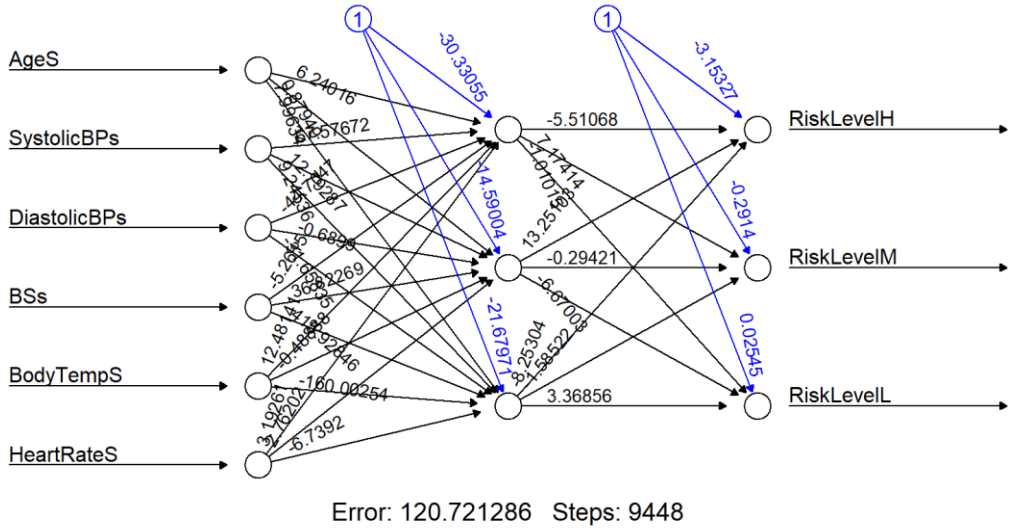
The maternal health risk dataset monitors pregnant women for the purpose of predicting maternal health risk during pregnancy. The data has been collected from different hospitals, community clinics, maternal health cares from rural areas in Bangladesh through a IoT based monitoring system. The response variable is Risk Level, which is the predicted risk intensity during pregnancy. The levels are high risk, medium risk, and low risk. There are six explanatory variables, which are Age, SystolicBP, DialosticBP, BS, BodyTemp, and HeartRate. Age is the age of the woman when pregnant. SystolicBP is the upper value of blood pressure in mmHg. DialosticBP is lower value of blood pressure in mmHg. BS is blood glucose levels in mmol/L. BodyTemp is the body temperature of the woman when pregnant. HeartRate is the heart rate of the woman when pregnant. This dataset was found on the UC Irvine Machine Learning Repository. The creator of this dataset is Marzia Ahmed from Daffodil International University.

Preprocessing Steps

Since all the explanatory variables are numeric with scales larger than 0 and 1, I scaled all of them, so that they have numbers between 0 and 1. The response variable, risk level, is categoric with categories high risk, medium risk, and low risk. I created dummy variables for each categoric. The created dummy variables are RiskLevelH for high risk, RiskLevelM for medium risk, and RiskLevelL for low risk.

Neural Network 1

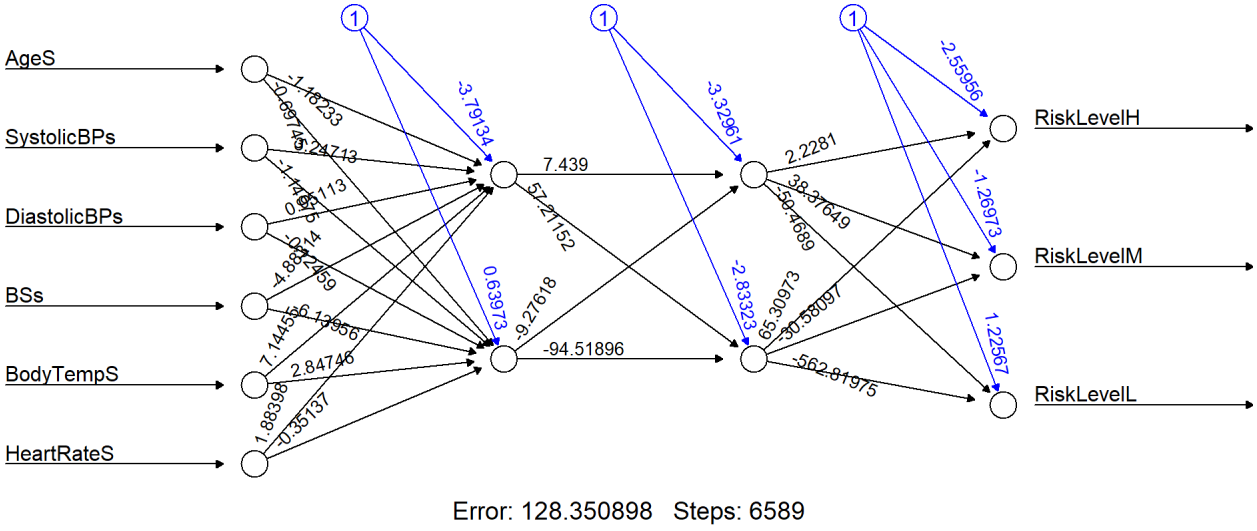
This is the first neural network model where the inputs are the six scaled explanatory variables of Age, SystolicBP, DialosticBP, BS, Bodytemp, and HeartRate. This model has one hidden layer with three nodes. Then there are three different outputs of RiskLevelH, RiskLevelM, and RiskLevelL because the response variable is categoric



Neural Network 1 Plot

Neural Network 2

This is the neural network plot for the second neural network generated. This neural network also has the six scaled explanatory variables as the inputs. This neural network has two hidden nodes with two nodes each. This neural network also has three outputs since the response variable is categoric.



Neural Network 2 Plot

Comparing The Neural Networks

To evaluate the two neural networks, a confusion matrix was generated for both neural network models on the validation dataset.

Below is the confusion matrix for neural network 1. The model had 65% accuracy. 0 corresponds with high risk, 1 with medium risk, and 2 with low risk. The model had 89 accurate predictions of high risk, but wrongfully predicted 17 times that it was high risk when it really was medium risk. It had 47 accurate predictions of medium risk, but wrongfully predicted 6 times that it was medium risk when it really was high risk, and 22 times wrongly predicted that it was medium risk when it really was low risk. There were 131 correct predictions of low risk, but 17 wrong predictions where it really was high risk and 77 wrong predictions where the right answer was medium risk.

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2
         0  89  17   0
         1   6  47  22
         2  17  77 131

Overall Statistics

               Accuracy : 0.6576
                 95% CI : (0.6092, 0.7037)
    No Information Rate : 0.3768
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.4773

 Mcnemar's Test P-Value : 2.006e-11
```

Below is the confusion matrix for neural network 2. This model had 68% accuracy. It had 89 accurate predictions of high risk, but 13 wrong predictions where it predicted high risk when it really was medium risk. There were 48 accurate predictions of medium risk, but 6 wrong predictions of medium risk when it was high risk and 10 wrong predictions of medium risk when it was low risk. There were 143 correct predictions of low risk, but 17 wrong predictions of low risk when it really was high risk and 80 wrong predictions of low risk when it really was medium risk.

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2
         0  89  13   0
         1   6  48  10
         2  17  80 143

Overall Statistics

               Accuracy : 0.6897
                 95% CI : (0.6422, 0.7344)
    No Information Rate : 0.3768
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.5249

 Mcnemar's Test P-Value : 5.867e-16

Statistics by Class:
```

Which Model Performed Better?

Neural network 2 performs better than neural network 1. When looking at the confusion matrix, neural network 2 had a higher accuracy with 68% than neural network 1, which had 65% accuracy. Neural network 2 had more true positives than neural network 1.