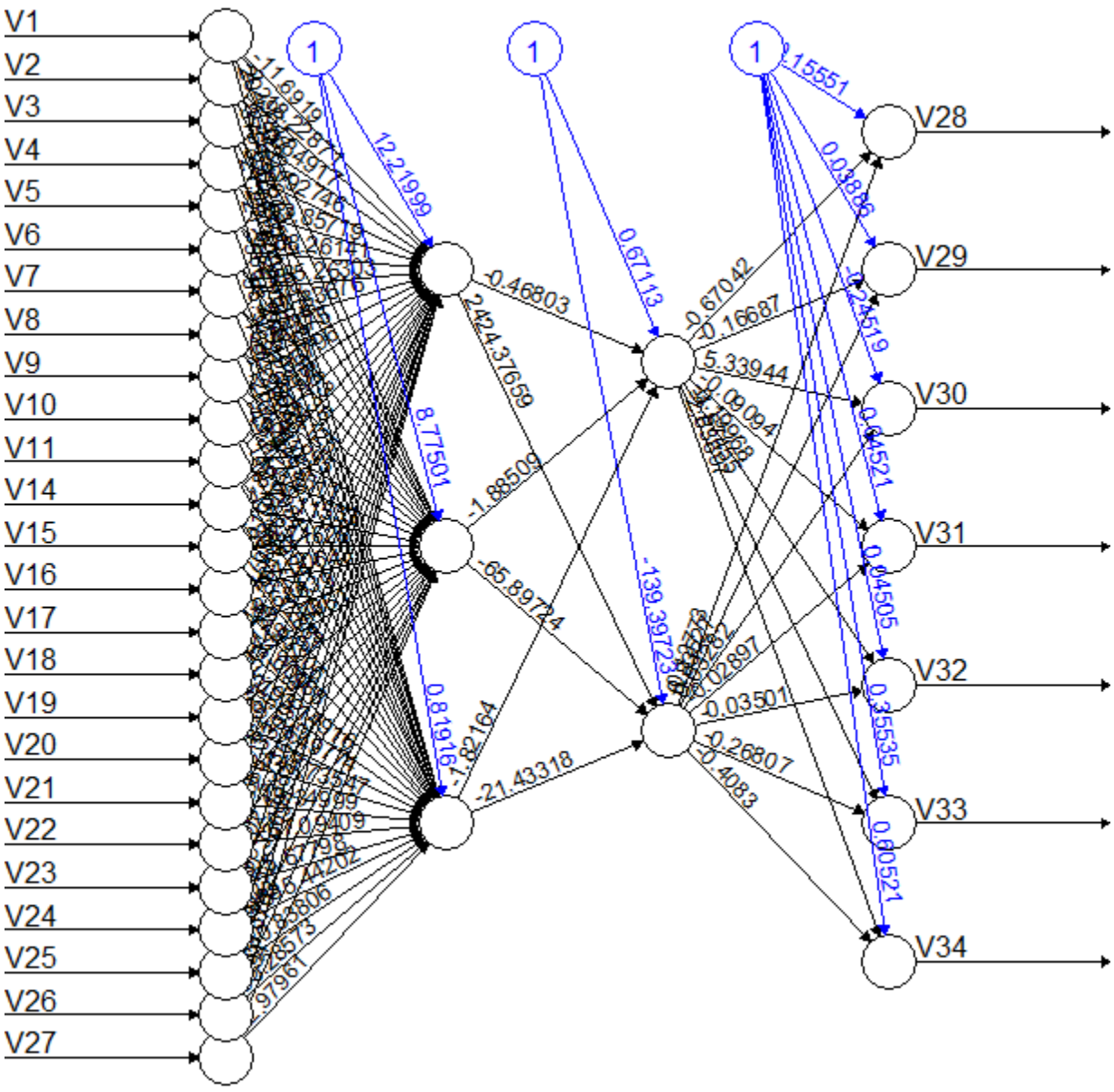
Assignment 11

Steel Plates Faults Data Set

Use Neural Network for steel plates fault prediction:

```
> # Read data using fread from the given URL
> still_data = fread("https://archive.ics.uci.edu/ml/machine-learning-databas
es/00198/Faults.NNA")


> data=still_data[,c(-12,-13)]
> # Extract max and min value from each column for scaling
> max=apply(data, 2, max)
> min=apply(data, 2, min)

# Scale only contineous variables from the dataset and exclude categorical va
riables for scaling

> scale_data = data[,1:25]

# Extract max and min value from each column for scaling
> max=apply(scale_data, 2, max)
> min=apply(scale_data, 2, min)

> scaled = as.data.frame(scale(scale_data,center = min,scale = max-min))

> View(scaled)

#lets predict all the 7 variable in one model for the still dataset

#bind the  7 variable column with the scaled data


> all_data <- cbind(scaled,data[,26:32])

> View(Zscratch_data)

> index = sample(1:nrow(all _data),round(0.75*nrow(all_data)))

> train = all_data[index,]
> test = all_data[-index,]

# Create formula for NN using the below function
> names = names(Zscratch_data)

> f = as.formula(paste("V28 + V29 + V30 + V31 + V32 + V33 + V34 ~ ",paste(nam
es[!names %in% c("V28","V29","V30","V31","V32","V33","V34")], collapse = "+")
))
> f
V28 + V29 + V30 + V31 + V32 + V33 + V34 ~ V1 + V2 + V3 + V4 +
    V5 + V6 + V7 + V8 + V9 + V10 + V11 + V14 + V15 + V16 + V17 +
    V18 + V19 + V20 + V21 + V22 + V23 + V24 + V25 + V26 + V27

# Neural Net for layer (3,2)
> nn = neuralnet(f,data=train, hidden=c(3,2), linear.output = T,act.fct = "lo
gistic",stepmax = 1000000)


> plot(nn)
```




```
# Compute the values of 7 predictor variable using test data
> test_nn = compute(nn,test[,1:25])
# Convert computed results to categorical values (0,1)
> test_cat = ifelse(test_nn$net.result > 0.5,1,0)

> library(caret)
> confusionMatrix(test_cat[,1],test$V28)
> confusionMatrix(test_cat[,2],test$V29)
> confusionMatrix(test_cat[,3],test$V30)
> confusionMatrix(test_cat[,4],test$V31)
> confusionMatrix(test_cat[,5],test$V32)
> confusionMatrix(test_cat[,6],test$V33)
```

Confusion Matrix result for each predictor variable

```
> confusionMatrix(test_cat[,1],test$V28)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 456  29
         1   0   0

               Accuracy : 0.9402062
```

We can see that misclassification is very high, in fact no class 1 predicted. Thus V28 (Pastry) is not a good attribute to predict fault.

```
> confusionMatrix(test_cat[,2],test$V29)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 428  12
         1  19  26

               Accuracy : 0.9360825
```

Accuracy is good here and misclassification is comparatively low, thus V29 (Z_Scratch) can be a good attribute to predict fault.

```
> confusionMatrix(test_cat[,3],test$V30)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 372   5
         1   5 103

               Accuracy : 0.9793814
```

Accuracy is good here and misclassification is very low, thus V30 (K_Scatch) must be a good attr ibute to predict fault.

```
> confusionMatrix(test_cat[,4],test$V31)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 464  21
         1   0   0

               Accuracy : 0.956701
```

We can see that misclassification is very high, in fact no class 1 predicted. Thus V31 (Stains) is not a good attribute to predict fault.

```
> confusionMatrix(test_cat[,5],test$V32)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 467  18
         1   0   0

               Accuracy : 0.9628866
```

We can see that misclassification is very high, in fact no class 1 predicted. Thus V32 (Dirtiness) is not a good attribute to predict fault.

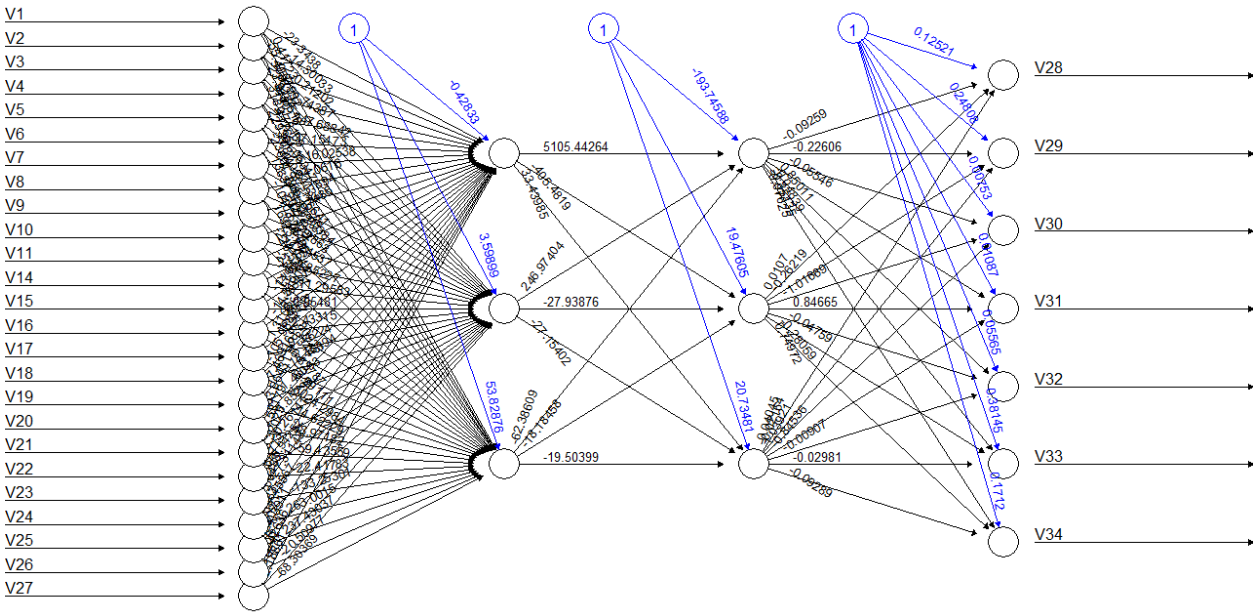
```
> confusionMatrix(test_cat[,6],test$V33)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 376 109
         1   0   0

               Accuracy : 0.7752577
```

We can see that misclassification is very high, in fact no class 1 predicted. Thus V33 (Bumps) is not a good attribute to predict fault.

```
> confusionMatrix(test_cat[,7],test$V34)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 323 162
         1   0   0

               Accuracy : 0.6659794
```

We can see that misclassification is very high, in fact no class 1 predicted. Thus V33 (Other_Faults) is not a good attribute to predict fault.

```
# Run NN for Hidden layer (4,2)

> nn4 = neuralnet(f,data=train, hidden=c(4,2), linear.output = T,act.fct = "logistic")
Warning message:
algorithm did not converge in 1 of 1 repetition(s) within the stepmax

# Reducing the Hidden layer (3,3)

> nn4 = neuralnet(f,data=train, hidden=c(3,3), linear.output = T,act.fct = "logistic",stepmax = 10000000)

> plot(nn4)
```




```
> confusionMatrix(test_cat4[,1],test$V28)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 456  29
         1   0   0

               Accuracy : 0.9402062

> confusionMatrix(test_cat4[,2],test$V29)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 447  38
         1   0   0


> confusionMatrix(test_cat4[,3],test$V30)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 373  13
         1   4  95

               Accuracy : 0.9649485


> confusionMatrix(test_cat4[,4],test$V31)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 458   3
         1   6  18

               Accuracy : 0.9814433
```

```
> confusionMatrix(test_cat4[,5],test$V32)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 467  18
         1   0   0

               Accuracy : 0.9628866

> confusionMatrix(test_cat4[,6],test$V33)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 376 109
         1   0   0

               Accuracy : 0.7752577

> confusionMatrix(test_cat4[,7],test$V34)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 265  63
         1  58  99

               Accuracy : 0.7505155
```

Thus, for (3,3) hidden layer, V30 and V31 are the good attribute to predict t
he fault.