26 BE 7082 + 26 PH 7028 + 20 BME 7082

Introduction to Data Science

Autumn 2020

MB Rao

Homework No. 9 Due date: December 01, 2020 Maximum points: 30

Theme: Neural Networks – Understand the mechanics behind the models

Download the data ‘bodyfat’ from the ‘TH.data’ package. Build a model for predicting body fat. Determination of body fat is complex. We have data from 71 women on DEXfat (body fat) + some anthropometric measurements + age.

1. Ignore the last four columns of ‘bodyfat.’ Create a new folder. Obtain summary statistics. 1 + 1 points

df<-subset(bodyfat, select=c(age, DEXfat, waistcirc, hipcirc, elbowbreadth, kneebreadth))

head(df)

summary(df)

age DEXfat waistcirc

Min. :19.00 Min. :11.21 Min. : 65.00

1st Qu.:42.00 1st Qu.:22.32 1st Qu.: 78.50

Median :56.00 Median :29.63 Median : 85.00

Mean :50.86 Mean :30.78 Mean : 87.38

3rd Qu.:62.00 3rd Qu.:39.33 3rd Qu.: 99.75

Max. :67.00 Max. :62.02 Max. :117.00

hipcirc elbowbreadth kneebreadth

Min. : 88.00 Min. :5.200 Min. : 7.200

1st Qu.: 96.75 1st Qu.:6.200 1st Qu.: 8.600

Median :103.00 Median :6.500 Median : 9.200

Mean :105.28 Mean :6.508 Mean : 9.301

3rd Qu.:111.15 3rd Qu.:6.900 3rd Qu.: 9.800

Max. :132.00 Max. :7.400 Max. :11.800

2. Fit a multiple regression model with DEXfat as the response variable. Check normality and homoscedasticity. Write the prediction equation. Identify the significant predictors. Calculate the MSE (Mean Square Error).

1 + 1 + 1 + 1 + 1 points

> index<-sample(1:nrow(df), round(0.75\*nrow(df)))

> head(index)

[1] 70 5 22 54 6 11

> train<-df[index,]

> test<-df[-index, ]

> #multiple regression

> MR<-lm(DEXfat~., data=train)

> summary(MR)

Call:

lm(formula = DEXfat ~ ., data = train)

Residuals:

Min 1Q Median 3Q Max

-8.8423 -2.3994 -0.1066 2.4301 11.1969

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -57.41561 10.68189 -5.375 2.34e-06 \*\*\*

age 0.04768 0.04779 0.998 0.323526

waistcirc 0.28830 0.09292 3.103 0.003241 \*\*

hipcirc 0.41249 0.11529 3.578 0.000816 \*\*\*

elbowbreadth -0.75327 1.42892 -0.527 0.600561

kneebreadth 2.38901 1.09574 2.180 0.034280 \*

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

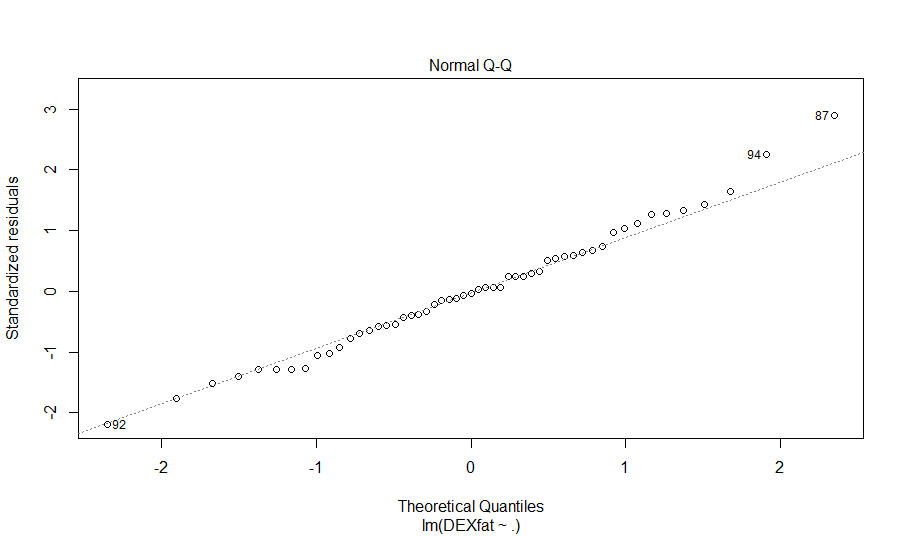
Residual standard error: 4.285 on 47 degrees of freedom

Multiple R-squared: 0.8607, Adjusted R-squared: 0.8459

F-statistic: 58.1 on 5 and 47 DF, p-value: < 2.2e-16

**Check for normality**

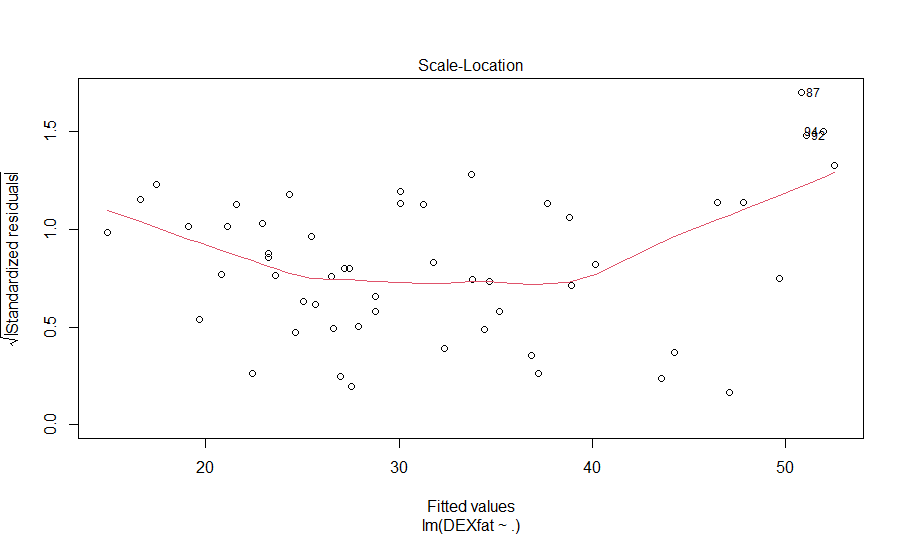
plot(MR, 2)



Since points lie approximately on the reference line we can assume normality.

**Check for homoscedasticity**

plot(MR, 3)



Since the residuals are spread evenly around the separation line, we can assume homogeneity of variance,

Prediction Equation:

**Significant Predictors:**

**Waistcirc, HipCirc, kneebreadth**

> #MSE

> pred<-predict(MR, newdata=test)

> head(pred)

47 50 65 67 70 75

37.43982 23.29689 37.35320 34.53567 27.07475 47.64319

**> MSEmr<-sum((pred - test$DEXfat)^2)/nrow(test)**

**> MSEmr**

**[1] 26.88094**

3. Fit a neural network model with three nodes in the hidden layer. Make sure that each variable is transformed to lie in between zero and one. What is the total number of weights and biases in the model? Write the prediction model. Calculate the MSE. Produce the plot. How many steps were taken to get the final model? 1 + 1 + 1 + 1 + 1 + 1 + 1 points

#step1:max and min column wise

Maxs<-apply(df, 2, max)

Maxs

Mins<-apply(df, 2, min)

Mins

Scaleddf<-scale(df, center=Mins, scale=Maxs-Mins)

class(Scaleddf)

head(Scaleddf)

#step2, select train set and test set

head(index)

trainset<-Scaleddf[index, ]

testset<-Scaleddf[-index, ]

dim(trainset)

dim(testset)

#neural network with three hidden nodes

Neural1<-neuralnet(DEXfat~., data=trainset, hidden=3, linear.output=TRUE)

Neural1

plot(Neural1)

Pred1<-predict(Neural1, newdata=testset)

Pred1Orig<-(Pred1)\* (max(df$DEXfat)-min(df$DEXfat))+min(df$DEXfat)

testDEXfat1<-(testset[, 2])\* (max(df$DEXfat)-min(df$DEXfat))+min(df$DEXfat)

MSENEural1<-sum((Pred1Orig-testDEXfat1)^2)/18

**MSENEural1**

**9.50**

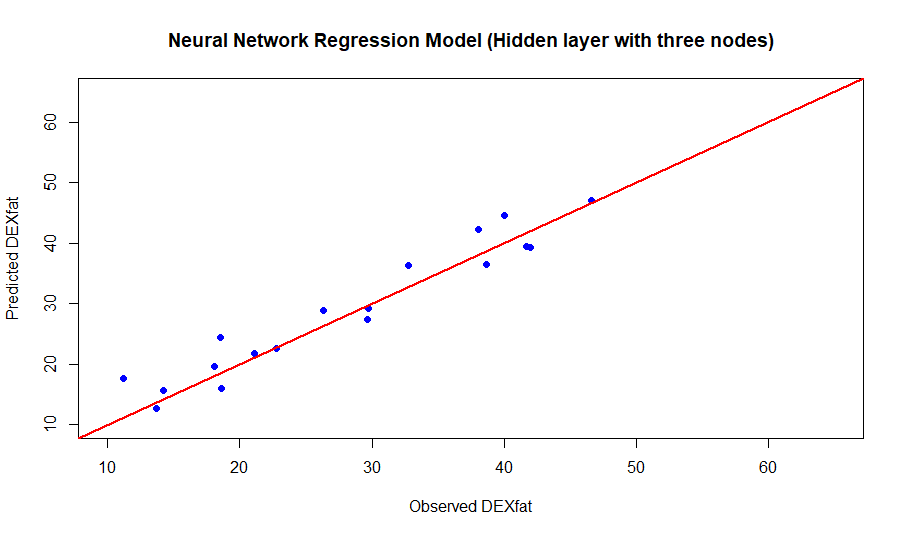
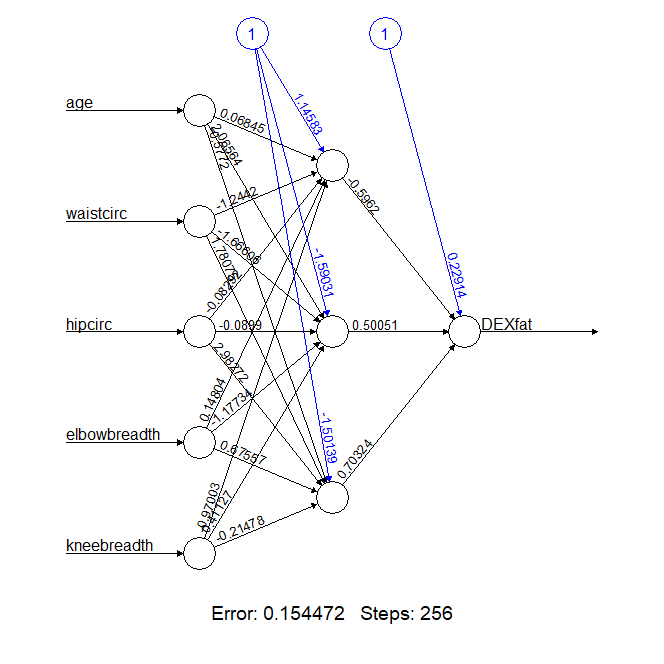
**Total number of weights: 18**

**Total number of biases =4**

**Prediction Model:**

**DEXfat=0.23-0.60\*(1/(1+exp(0.07\*age-1.24\*waistcirc-0.08\*hipcirc+0.15\*elbowbreadth+0.97\*kneebreadth+1.15)))+0.5\*(1/(1+exp(2.07\*age—1.67\*waistcirc-0.09\*hipcirc-1.18\*elbowbreadth+0.41\*kneebreadth-1.60)))+0.70\*(1/(1+exp(0.58\*age+1.78\*waistcirc+2.98\*hipcirc+0.68\*elbowbreadth-0.21\*kneebreadth-1.50)))**

**Steps Taken:** **256**



4. Fit a neural network model with four nodes in the hidden layer. Make sure that each variable is transformed to lie in between zero and one. What is the total number of weights and biases in the model? Write the prediction model. Calculate the MSE. Produce the plot. How many steps were taken to get the final model? 1 + 1 + 1 + 1 + 1 + 1 + 1 points

Neural2<-neuralnet(DEXfat~., data=trainset, hidden=4, linear.output=TRUE)

Neural2

plot(Neural2)

Pred2<-predict(Neural2, newdata=testset)

Pred2

Pred2Orig<-(Pred2)\* (max(df$DEXfat)-min(df$DEXfat))+min(df$DEXfat)

(Pred2Orig)

testDEXfat2<-(testset[, 2])\* (max(df$DEXfat)-min(df$DEXfat))+min(df$DEXfat)

(testDEXfat2)

MSENEural2<-sum((Pred2Orig-testDEXfat2)^2)/18

**MSENEural2**

**9.33**

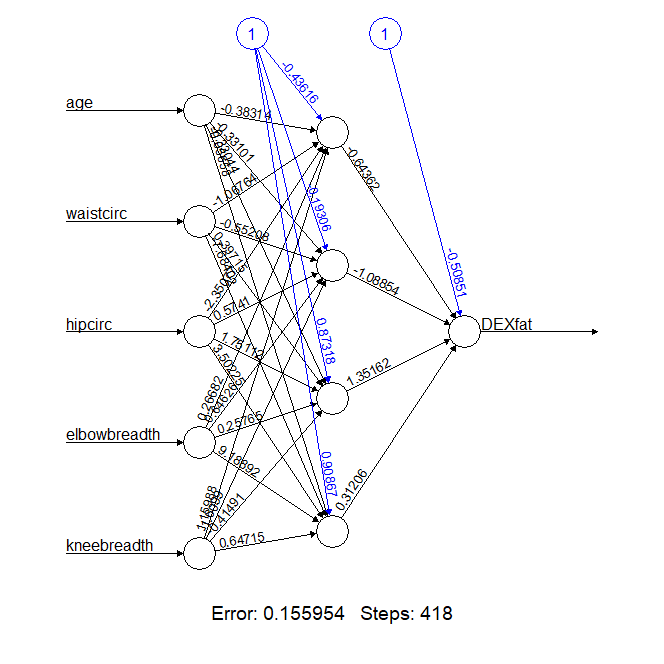
**Prediction Model:**

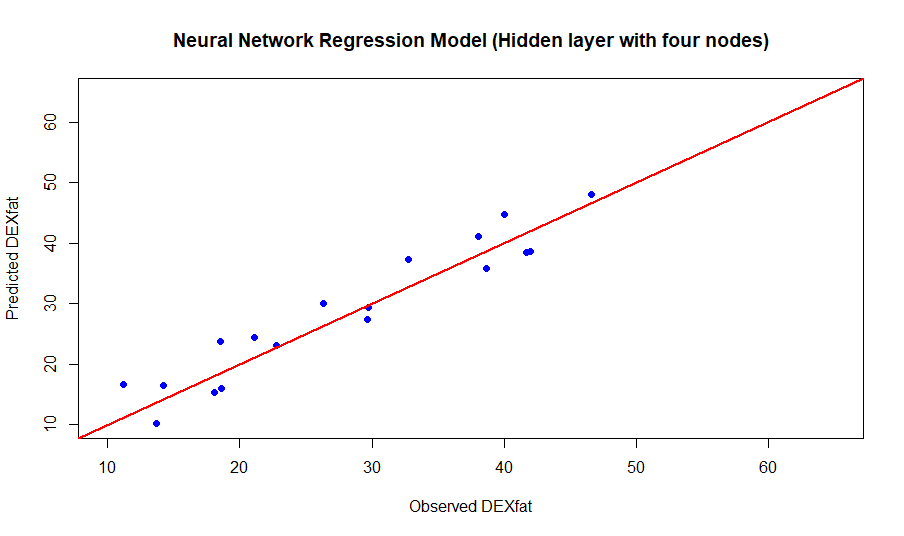
**DEXfat= - 0.51 – 0.64\*(1/(1+exp(-0.38\*age-1.07\*waistcirc-2.35\*hipcirc+0.27\*elbowbreadth+1.16\*kneebreadth-0.44)))-1.09\*(1/(1+exp(-0.33\*age-0.55\*waistcirc+0.57\*hipcirc+0.64\*elbowbreadth+1.60\*kneebreadth-0.19)))+1.35\*(1/(1+exp(0.33\*age+0.39\*waistcirc+1.75\*hipcirc+0.26\*elbowbreadth-0.41\*kneebreadth+0.87))) +0.31 \*(1/(1+exp(-0.05\*age+7.69\*waistcirc+3.50\*hipcirc+9.18\*elbowbreadth-0.65\*kneebreadth+0.91)))**

**Steps Taken: 418**

**Total number of weights: No. of Inputs\*Number of Hidden layers + Number of Hidden Layers = 24**

**Total number of biases =5**





5. Fit a neural network model with five nodes in the hidden layer. Make sure that each variable is transformed to lie in between zero and one. What is the total number of weights and biases in the model? Write the prediction model. Calculate the MSE. Produce the plot. How many steps were taken to get the final model? 1 + 1 + 1 + 1 + 1 + 1 + 1 points

#neural network with five hidden nodes

Neural3<-neuralnet(DEXfat~., data=trainset, hidden=5, linear.output=TRUE)

Neural3

plot(Neural3)

Pred3<-compute(Neural3, testset[,1:6])

Pred3$net.result

Pred3Orig<-(Pred3$net.result)\*(max(df$DEXfat)-min(df$DEXfat))+min(df$DEXfat)

(Pred3Orig)

testDEXfat3<-(testset[, 2])\*(max(df$DEXfat)-min(df$DEXfat))+min(df$DEXfat)

(testDEXfat3)

MSENEural3<-sum((Pred3Orig-testDEXfat3)^2)/18

> **MSENEural3**

**[1] 8.869208**

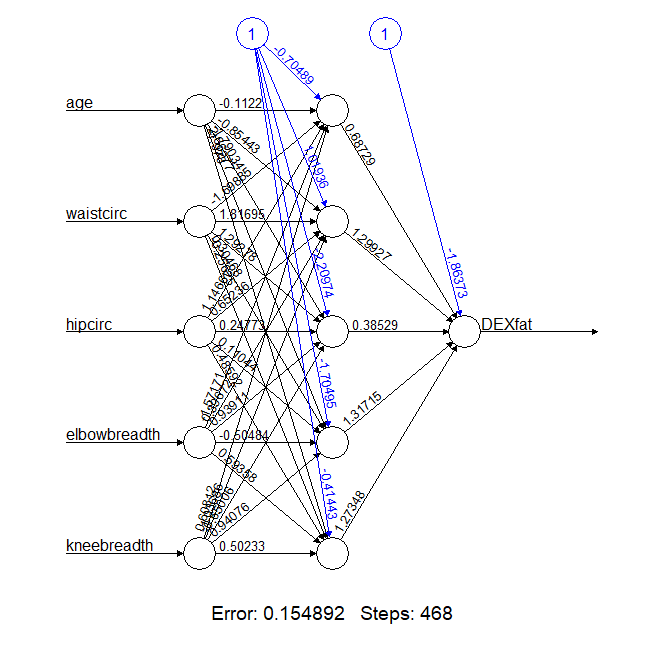
**Prediction Model:**

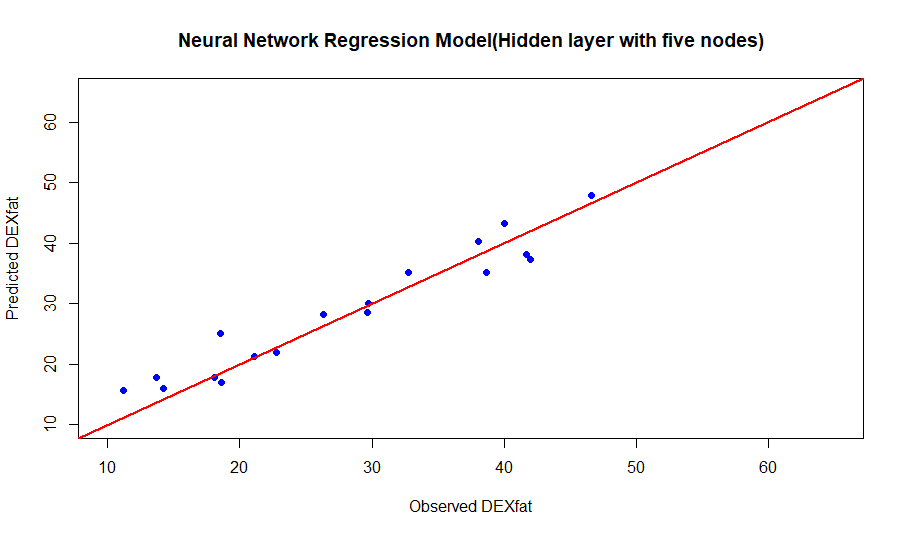
**DEXfat= - 1.86 + 0.68\*(1/(1+exp(-0.11\*age-1.69\*waistcirc+1.16\*hipcirc+1.57\*elbowbreadth+0.61\*kneebreadth-0.70)))-1.30\*(1/(1+exp(-0.85\*age-1.82\*waistcirc+0.65\*hipcirc+0.39\*elbowbreadth+1.60\*kneebreadth+1.02)))+0.39\*(1/(1+exp(1.80\*age+1.30\*waistcirc+0.25\*hipcirc+0.94\*elbowbreadth-0.65\*kneebreadth+1.12))) +1.32 \*(1/(1+exp(-0.25\*age+0.3\*waistcirc+0.11\*hipcirc-0.50\*elbowbreadth-0.94\*kneebreadth-1.70)))+ 1.27 \*(1/(1+exp(-0.51\*age-0.25\*waistcirc+0.49\*hipcirc-0.60\*elbowbreadth-0.50\*kneebreadth-0.41)))**

**Steps Taken: 468**

**Total number of weights: No. of Inputs\*Number of Hidden layers + Number of Hidden Layers = 30**

**Total number of biases = 6**





6. Contrast the MSEs from 2, 3, 4, and 5. Comment. 2 points

The MSEs obtained from neural networks are significantly lower than the MSE obtained using Multiple regression. Furthermore, MSE reduces as we increase number of nodes in the hidden layer. For Multiple regression I had 5 parameters, for NN with 3 nodes in the hidden layer I had 22 parameters, for NN with 4 nodes in the hidden layer I had 29 parameters and for NN with 5 nodes in the hidden layer I had 36 parameters. As a result, NN models are approximately 3 times more accurate (24% compared to 8% MSE) compared to Multiple regression.