Introduction to Machine Learning - Lab 8

Gustav Sternelöv

Friday, December 11, 2015

Assignment 1

1.1

A function that uses the simple perceptron algorithm is implemented with the following code:

```
### Assignment 1
## 1.1
# "lambda" is the learning rate, "until" the parameter for the stop condition
spl <- function(x, y, lambda, weights, until){
  yHat <- 0
  MisClass1 <- 0
  MisClass2 <- 1
  # Loop that runs until stopping condition is met.
  while(abs(MisClass2 - MisClass1) > until){
    # The misclass value for the latest iteration is saved for comparison with
    # the next misclass value
    MisClass2 <- MisClass1
    for(i in 1:nrow(x)){
      # The weights are multiplied with i:th row
      linComb <- sum(weights * x[i, ])
      # Uses the sign function to classify the observation
      yHat[i] <- sign(linComb)
      # Updates the weights
      weights <- weights + 0.1*(y[i] - yHat[i]) * x[i, ]
    }
    # Calculates the misclassification value
    confMat <- table(yHat, y)
    MisClass1 <- 1 - sum(diag(confMat)) / sum(confMat)
  }
  RetU <- list(Wt = weights, miscTr = MisClass1)
  return(RetU)
}
# The classify the test set with the obtained model the following code is used
testSPL <- spl(x=inputs, y=train[,58], lambda=0.1, weights=init_weights, until=0.01)
yHatTest <- 0
for(i in 1:nrow(testInput)){
  # Uses the weights returned by the function on every row
  linCombTest <- sum(testSPL$Wt * testInput[i,])
  yHatTest[i] <- sign(linCombTest)
}
confMatTest <- table(yHatTest, test$Spam)
MisClass3 <- 1 - sum(diag(confMatTest)) / sum(confMatTest)
```

1.2

Data is divided into a train and a test data set. The train set consists of 70 % of the original data set and the test set of the remaining 30 %.

1.3

The initial weights are obtained by simulating from the normal distribution. These weights are used together with the train data set to fit a simple perceptron model. Regarding the other parameters the learning rate λ is set to *0.1* and the stopping condition is set to *0.01*.

1.4

The fitted model in *1.3* is used to classify the observations in the test set. Then, the misclassification rate is computed to be 11.59 %. With such a low percentage of misclassified observations the fitted model seem to be a good fit to this data.

1.5

Step *1.3* and *1.4* is repeated, but with another seed. This time the misclassification rate when classifying the test data is computed to be 8.69 %. Then, again, the conclusion is that the fitted model is a good fit to this data set.

1.6

For the variable *Spam* the class *-1* is changed to *0*. Then a logistic regression is fitted where *Spam* is the target variable. The model is fitted on train data and evaluated by predicting the class for the observations in the test set. A confusion matrix of the predicted values against the observed values is created and can be seen below. Since 17 of the 138 observations is misclassified the obtained misclassification rate is 12.3 %.

```
##          fitted
## test.Spam  0  1
##         0 71 10
##         1  7 50
```

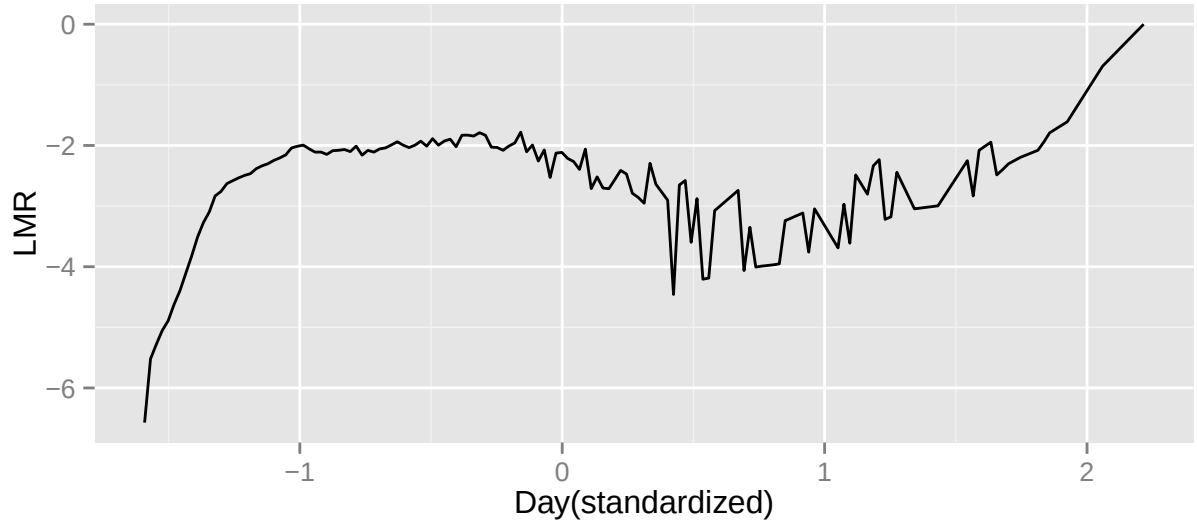
1.7

The two first models are identical simple perceptron models apart from the chosen seed. A comparison of these models gives that the second one are a little bit better since it has a lower misclassification rate for the test set (8.59 % vs 11.59 %). The third model was logistic regression model and with this model a misclassification rate of 12.3 % was obtained. That is the highest misclassification rate of all the fitted models. The simple perceptron models seem to be better fits, and of the three fitted models the second simple perceptron model gave the best result.

Assignment 2

2.1-2.2

The studied data set contains information about the mortality rate for fruit flies for each day. The data comes from a study where the theory that the mortality rates (probability of dying per unit time) of many organisms increase at an exponential rate was tested. The variable *LMR*, that is the logarithm of the variable *Rate*, is plotted against the variable *Day*(standardized).



The relationship between *LMR* and *Day* is quite different for different periods. During the first days the mortality rate is increasing rapidly. Then, during the middle of the observed time period, the mortality rate is constant a short while before it decreases a little. At the end of the observed time period the mortality rate increases, not as fast as in the beginning but still rather fast.

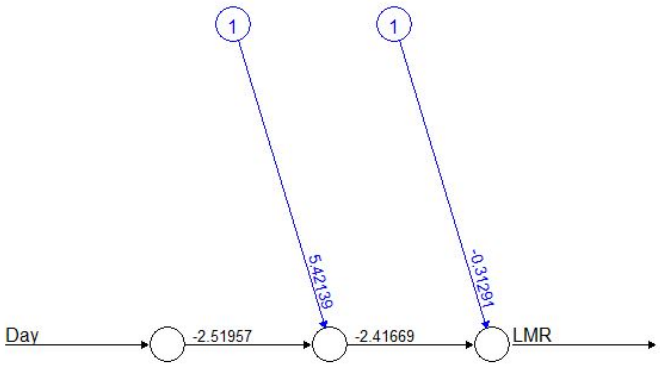
2.3

A multilayer perceptron network with a single hidden neuron, *tanh* as the activation function ,a maximum number of iterations that is $1 * 10^6$ and a treshold for stopping the fitting of the model at *0.1*.

a)

The fitted formula of the network:

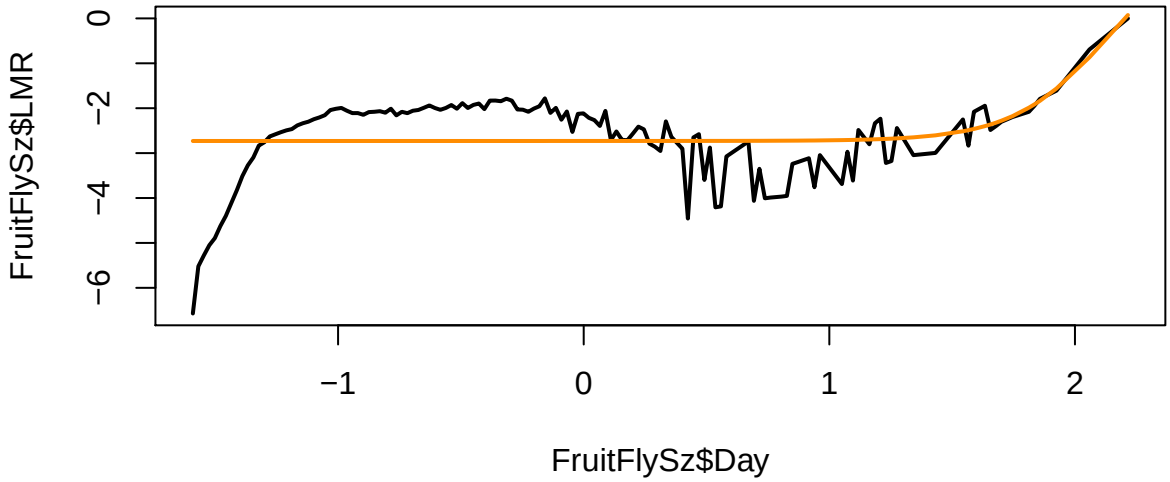
A plot visualizing the network's weights.



Error: 49.703647 Steps: 607

b)

The fitted values obtained are visualized with the orange line and compared against the observed values (the black line).



It can be concluded that this model not is an especially good fit. Problaby this model is too simple and a little more complex model could give better results.

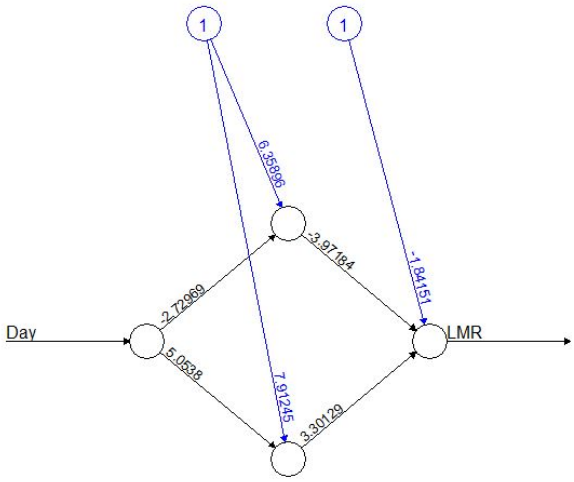
2.4

Again, a multilayer perceptron model is fitted. The new model has the same settings as the model in *2.3*, except that this model has two hidden neurons instead of one.

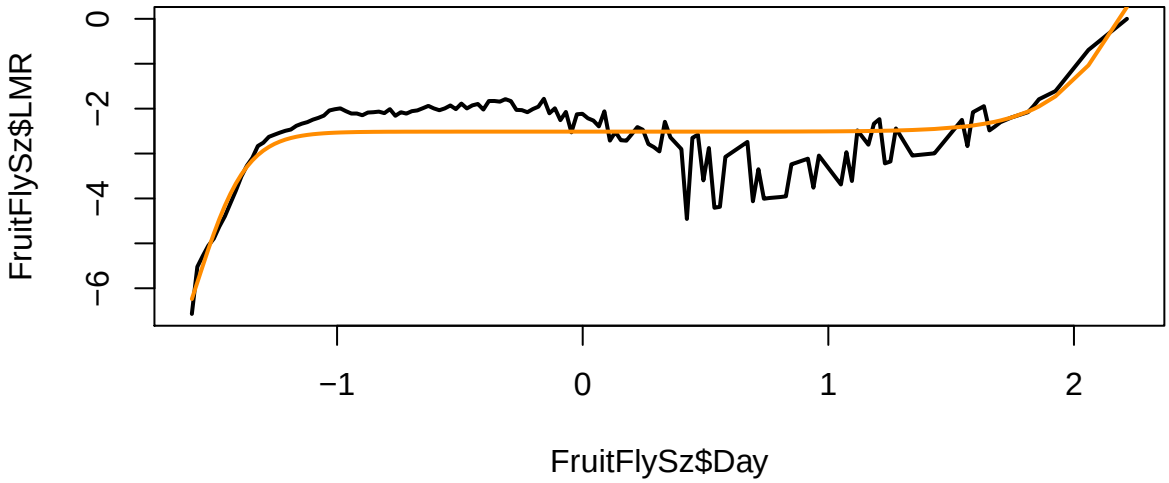
a)

The fitted formula of the network:

A plot visualizing the network's weights.



b)

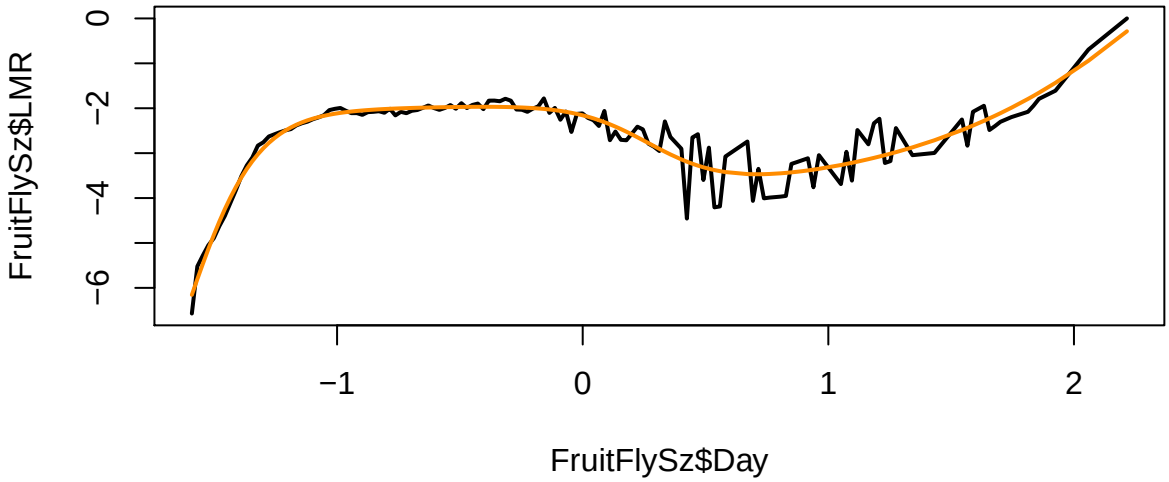


Compared to the fitted values for the model in *2.3*, the new model with two hidden neurons seem to be better. The use of two hidden neurons makes the model more flexible than before. That is why more of the general pattern in the data is captured with the new model. However, even if these model is better than the former model there is still some room for improvement.

2.5

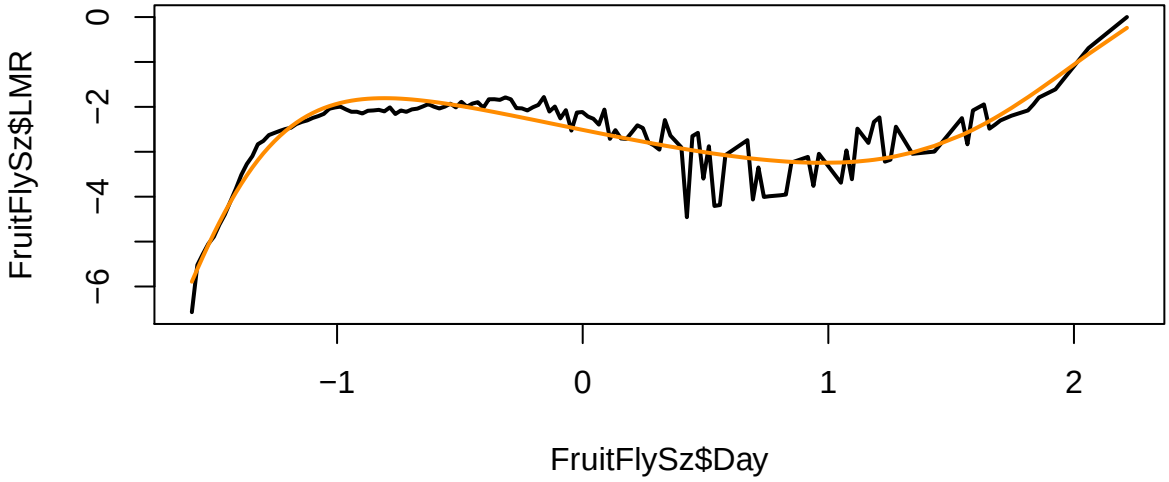
Four hidden neurons

Error:6.21 Number of iterations:2355



Five hidden neurons

Error:9.29 Number of iterations:8201



Comparison of the models

Both models are good fits to the data set, especially the first model with four hidden neurons seem to be a really good fit. This visual conclusion is also confirmed by the number of iterations and the error which is lower for the model with four hidden neurons.

When comparing the model in *2.4* with two hidden neurons and the model with four hidden neurons it is evident that a more complex model is a better fit to this data. However, when a even more complex model with five hidden neurons is fitted it does not result in a better fit.

It seems like a complex model is needed, but not a too complex model. A too simple model lacks flexibility and is therefore not a good fit. A too complex model