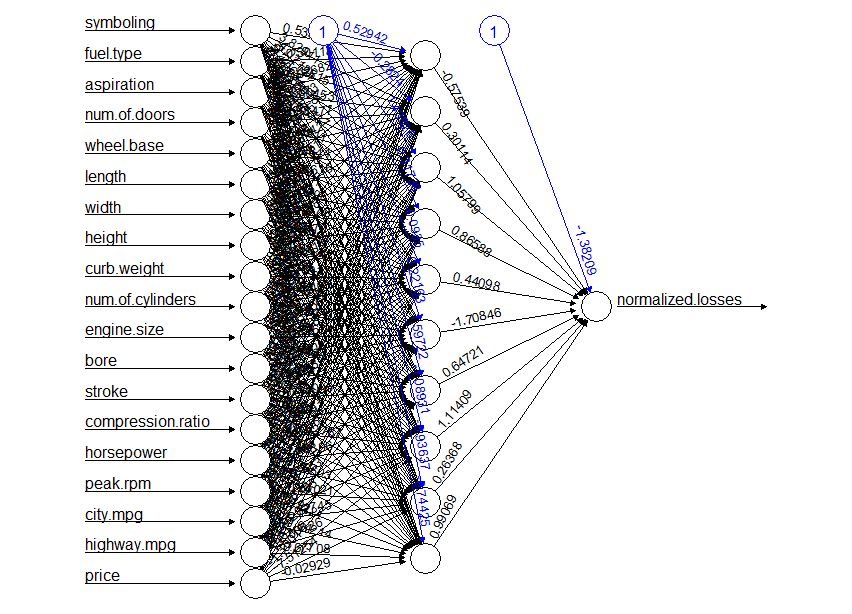
Project forum 7

Neural nets:

I’m working with automobile data used by insurance companies. I predicted the normalized losses (the third factor is the relative average loss payment per insured vehicle year) based on 19 different numerical factors as can be seen in the neural net plot below. Some categorical variables could not be used but I converted the factor variables with 2 levels into 1’s and 2’s since (though this can’t be done with factors with more than 2 levels because it would imply some sort of relationship between the different levels).

I has 159 rows and used 75% for training data and 25% for testing. I set the seed to 12345 to get repeatable data and set hidden to 10. The correlation between the predicted and real values for normalized losses was 0.6951027657. I tried different seed and hidden values and they showed no trend. Sometimes bigger values gave worse results and sometimes smaller values gave worse results. The correlation went as low as 0.2 in some seed values. Because of this. I don’t think that neural nets are very good for this dataset.

When I had done regression on this earlier, a lot of the most significant dependent variables were related to the make (i.e. brand of automobile e.g. Chevrolet) which is a categorical variable and cannot be used for neural nets.



Support Vector:

I classified my target variable (normalised losses) into five different categories—very low, low, medium, high, very high. Again, I has 159 rows and used 75% for training data and 25% for testing. I tried out a few different kernels and anovadot worked best for me. I had a training error rate of 0.09. I had 0.67% correctly predicted data which is similar to the result I had in neural nets.

