Mittuniversitetet
MID SWEDEN UNIVERSITY

Task 2 – Using a Neural Network on a fuel consumption dataset

Introduction

Based on the lab tutorial in neural networks, the task is to implement your own estimation system using a neural network with Bayesian Regularization.

The dataset describes fuel consumption on bus routes, with several input parameters and a ground truth fuel consumption column. This is used to estimate the fuel consumption of a planned bus route, with some co-variance error. The case closely resembles the forest log pile case from the tutorial. You are allowed to set all the parameters yourself and to design the neural network as you see fit. You are also allowed to take the error assessment script from the tutorial.

Include the MATLAB script code, the training diagrams and the best-case co-variance error that you can achieve. You are free to implement the neural network as you see fit. You are allowed to use the tutorial example as a template.

Some things to think about when you do the task:

1. Identity numbers in features

The identity numbers of the bus drivers, vehicles etc are very random. For a Support Vector Machine or any other non-nn method this will severely degrade the estimation accuracy when using the id of the drivers. There is also a degrading effect on the accuracy when using neural networks, although not as severe. In order to achieve higher accuracy, you should think about a way to sort/reset id numbers so they correspond better to other features in the dataset.

One way would be to sort the id numbers according to the mean fuel consumption connected to that ID. The downside of that would be that you would have to retrain the network when a new ID (new staff, new bus etc) is introduced into the dataset. Can you think of a way to sort the data where constant retraining would not be necessary?

2. List of available features

```
short_matrix_no_id(:,1)       Number of bus line
short_matrix_no_id(:,2)       Day of the year
short_matrix_no_id(:,3)       Weekday
short_matrix_no_id(:,4)       Hour of the day
short_matrix_no_id(:,5)       Travel time in minutes
short_matrix_no_id(:,6)       Vehicle ID
short_matrix_no_id(:,7)       Vehicle class ID
short_matrix_no_id(:,8)       Driver ID
short_matrix_no_id(:,9)       Distance
short_matrix_no_id(:,10)      Average outside temperature
short_matrix_no_id(:,11)      Tyre diameter
short_matrix_no_id(:,12)      Tyre width
short_matrix_no_id(:,13)      Vehicle model year
short_matrix_no_id(:,14)      Odometer
short_matrix_no_id(:,15)      Fuel used (ground truth)
short_matrix_no_id(:,16)      Type of day
short_matrix_no_id(:,17)      Swedish holiday (yes/no)
short_matrix_no_id(:,18)      Engine power
short_matrix_no_id(:,19)      Days since last service
```

Column 15 is the ground truth to be trained against.

3. Examination and feedback

Hand in a report (3-5 pages) with figures of training results, network configuration and best case co-variance error. If you make some sort of comparison between different setups or methods, include results of all setups and methods. The ambition level is up to you, but at least one working neural network with attached co-variance error is required.

You will be getting feedback on your solutions and, if applicable, suggestions of improvements.