TBMI26 – Computer Assignment Report  
Supervised Learning

Deadline – February 12 2018

Author/-s:

In order to pass the assignment you will need to answer the following questions and upload the document to LISAM. You will also need to upload all code in .m-file format. If you meet the deadline we correct the report within one week after the deadline. Otherwise we give no guarantees when we have time.

1. **Give an overview of the data from a machine learning perspective. Consider if you need linear or non-linear classifiers etc.**

For dataset 1 a linear classifier is sufficient. For dataset 2 and 3 you need a non-linear classifier.

1. **Explain why the down sampling of the OCR data (done as pre-processing) result in a more robust feature representation. See** [**http://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of+Handwritten+Digits**](http://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of+Handwritten+Digits)

Since the noise is reduced by this, the features become more robust. It becomes easier to distinguish between features.

1. **Give a short summery of how you implemented the kNN algorithm.**

First the Euclidean distance between all features was calculated, these were then sorted. The we choses the most frequent used label in the k-nearest neighbors.

1. **Explain how you handle draws in kNN, e.g. with two classes (k = 2)?**

We choose the label of the closest neighbor.

1. **Explain how you selected the best k for each dataset using cross validation. Include the accuracy and images of your results for each dataset.**

We used cross validation with 2 folds. First the distance was calculated from fold1 to fold2, and then from fold 2 to fold 1. The average of the two errors was stored. We did this for 10 different k values and plotted the average error for each one. The k with the highest accuracy was chosen.

1. **Give a short summery of your backprop network implementations (single + multi). You do not need to derive the update rules.**
2. **Present the results from the backprop training and how you reached the accuracy criteria for each dataset. Motivate your choice of network for each dataset. Explain how you selected good values for the learning rate, iterations and number of hidden neurons. Include images of your best result for each dataset, including parameters etc.**
3. **Present the results, including images, of your example of a non-generalizable backprop solution. Explain why this example is non-generalizable.**
4. **Give a final discussion and conclusion where you explain the differences between the performances of the different classifiers. Pros and cons etc.**
5. **Do you think there is something that can improve the results? Pre-processing, algorithm-wise etc.**