Programming Exercise 2

Due on 23rd November

Pick a classification dataset from the LIBSVM repository:

`http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/`

Tasks:

1. Partition the dataset into a training set and a validation set. Run a decision tree algorithm on the training set to generate a decision tree. Validate the decision tree on the validation dataset and report the total classification error (i.e. 0/1 error). Repeat the experiment with a different partition of training set/validation set. Are the resulting decision trees different? Explain why.

2. Run SVM to train a classifier, using a radial basis function as a kernel function. Apply cross-validation to evaluate different combinations of values of the model parameters (box constraint C and kernel parameter γ). How sensitive is the cross-validation error to changes in C and γ? Choose the combination of C and γ that minimizes the cross-validation error, train the SVM on the entire dataset and report the total classification error.

3. Train a Multi-Layer perceptron using the cross-entropy loss with ℓ-2 regularization (weight decay penalty). In other words, the activation function equals the logistic function. Plot curves of the training and validation error as a function of the penalty strength. How do the curves behave? Explain why.

You have to delive the source code for each section. You are free to use the programming language/library of your choice.