Pattern Recognition
Lab week 4

Useful matlab functions:

- `repmat` — Replicate and tile an array.
- `randperm` — Random permutation.

Assignment 1: *Supervised Learning: LVQ1.*

In Nestor you will find the files data_lvq_A.mat and data_lvq_B.mat, each of which contains 100 two-dimensional feature vectors. They belong to class A and class B, respectively. Your task is to build a LVQ1 classifier using this training data.

1. First investigate this data set. How many prototypes per class are theoretically suitable in this data set?
2. Implement the LVQ1 algorithm, see slide 6 of the lecture on LVQ. Use a constant learning rate $\eta = 0.01$. Determine the winner by minimal squared Euclidean distance (i.e. comparable to KNN with k = 1). Consider the following cases:
   a) 1 prototype for class A and 1 prototype for class B,
   b) 1 prototype for class A and 2 prototypes for class B,
   c) 2 prototypes for class A and 1 prototype for class B,
   d) 2 prototypes for class A and 2 prototypes for class B.

   Use all data in the LVQ1 training process. After each epoch, determine the number of misclassified training examples. We define the training error E as the number of misclassified training examples divided by the total number of data points (200). Plot E as a function of the number of epochs and stop the training when E becomes approximately constant.

hint: Think about how to choose the initial prototype(s).

hint: You will get benefits from your K nearest neighbor implementation from week 3.

hint: The learning rate η is the key factor in LVQ. You will get bonus points if you observe and report how the behavior depends on the learning rate.

Assignment 2: *Cross validation*

For case 1(c), apply 10-fold (cross validation for the) estimation of the classification error. This means, divide the training set in 10 equal subsets. Use 9 subsets to train an LVQ1 classifier and then use the remaining 1 subset to test it and determine the classification error. Use each of the subsets as a test set once, with the other 9 subsets as training sets. Compute the mean of the 10 values of the classification error computed in this way.

Assignment 3: *Relevance LVQ.*

For case 1(c), implement the relevance LVQ algorithm with global relevances (i.e. relevances that apply to the whole feature space and not just to one prototype or a class). Use 0.5 as initial value of the global relevances λ_1 and λ_2 . Plot the training error E and the relevances of the two features as a function of the number of epochs and stop the training when E becomes approximately constant. Use 10-fold (cross validation for the) estimation of the classification error.

hint: Read the lecture slides and use an appropriate value of the relevance learning rate η_λ.

Requirement:

You should hand in a report that contains (at least):

- for all problems: your Matlab code and your analysis of the experiment result.
- for problem 1: one example learning curve for each of the cases 1(a) to 1(d) and a corresponding 2D plot which shows the LVQ1 prototypes (make your data points and prototypes distinguishable, especially if you hand in a gray-level report.)
- for problem 2: the value of the test error as determined by 10-fold estimation
- for problem 3: one example learning curve of the training error and the relevances and a corresponding 2D plot which shows the LVQ prototypes; the test error as determined by 10-fold estimation.