CNPG-765/CMPT- 465 Neural Networks and Learning Systems
Homework-5

"**Wisconsin Diagnostic Breast Cancer (WDBC)**" is a famous, very popular benchmark 2-class classification problem https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic) .

This dataset consists of 569 instances.

In a typical experiment, 400 instances are used for learning, and 169 instances are used for testing.

1. (**undergraduate students - 100/100, graduate students – 50/100**)
   1) Solve a pattern recognition problem with the "WDBC" dataset using MLP. You should use functions LearningMLP and TestingMLP2, and the data file wdbc_MLP_Learning_Testing_Data.mat (containing a learning set and a test set) which were shared with you.
   You need to make 5 learning/testing experiments increasing the number of hidden neurons (starting from 2) in a hidden layer and see how your classification rate changes depending on the number of hidden neurons.
   2) After you determine a network topology giving the best classification rate, repeat your experiment 5 times for this topology starting from the different random weights. Find average accuracy (average classification rate) for this topology.
   Limit the number of learning iterations by 5000 and use 0.1 as a tolerance threshold to stop your learning process.
   3) Present your results in a technical report (you may summarize all of them in a single table) and make a conclusion about the best topology of a network leading to the highest classification rate.

2. (**Required for graduate students (50/100) and extra credit (30 points) for undergraduate students)**

   1) Modify the LearningMLP.m function **by changing its stopping criterion from minimization of learning RMSE to minimization of a validation RMSE**.
   You will also need to add a new calling parameter to LearningMLP.m – a validation set.
   2) Divide a test set (Testing_wdbc_MLP) into two sets: move the first 85 samples from this set to a validation set, which should be used during the learning process, and the remaining 84 samples should remain in the test set and used for testing.
   3) Make 5 experiments with the best topology, which you determined in Task 1, and try to reach a lower classification error than was reached there – **your learning process should stop if your validation error becomes lower than your best testing error in Task 1**.
   4) Present your results in a technical report (you may summarize all of them in a single table) and make a conclusion about the results.

3. (**Extra credit for all students – 20 points**) Compare a classical learning process (Task 1) with a learning process with validation (Task 2). Is it possible to reach higher classification accuracy by running a learning process with validation? Is a learning process with validation shorter in terms of the number of iterations than a classical one? Justify your conclusions by data.

Turn in your report.