Lab report for 1 hidden layer artificial neural network

1. Parameter and general structure
   * The entire ANN has 21x17 input unit, 10 hidden units, and 3 output units. The number 17 is determined by the window size. Each input can be chosen from 20 known alphabets. One additional input corresponds to the bias. The final implementation (submitted) incorporates the momentum and weight decay.
   * ***Learning rate = 0.25***
   * ***Momentum = 0.7***
   * ***Weight decay parameter = 0.001;***
   * Both the hidden/output unit use sigmoid function and turns out to perform better than rectified linear
   * **Early Stopping:** Test set’s result are reported based on the epoch when tuning set achieve the highest result
   * **Weight initialization:** the weight of the connections of ANN during initializations is randomized from -1 to 1.
   * The weight decay + momentum backpropagation is according to the following equation:



1. **Result on the 128 protein:** 
   * The Neural Network’s best performance is around 62%. The epochs beyond 300 are computationally expensive so I did not report it. As the training set best accuracy in this graph is only 75%, it is not expected early stopping would occur. Either run over 1000 epochs or run on smaller samples to observe early stopping.
2. **Effect of Learning Rate:** All Learning rate screen are based on 21x17 inputs, 10 hidden units, 3 output units. The input file (trial2) is a truncated one (12 proteins instead of 128) for fast computation.
   * It is worth noticing that, when learning speed is high, training set’s accuracy reaches maximum within smallest number of epochs. An extremely slow learning rate would make the performance rather retard, as the network essentially did not change during each backpropagation.

Fig 1: Learning with a fast speed

Fig 2: Learning with a medium speed

Fig 3: Learning with a slow speed

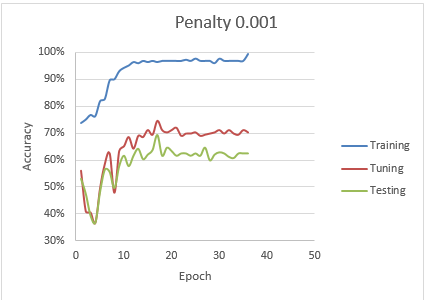
1. **Effect of Number of Hidden Units**: hidden units’ screen are based on 21x17 inputs, 10 hidden units, 3 output units. The input file (trial2) is a truncated one (12 proteins instead of 128) for fast computation.
   * My results implied that increasing the number of hidden units, would increase the instability of the training model. Considering the number of inputs are small, too many hidden units would incur point-to-point mapping, overfitting and pick up noise.

Fig 4: few hidden units

Fig 5: medium hidden units

Fig 6: lots of hidden unit

1. **Effect of weight decay parameter**: 2 screens were done, shown in Fig 7, 8. Weight decay’s screen are based on 21x17 inputs, 10 hidden units, 3 output units. The input file (trial2) is a truncated one (12 proteins instead of 128) for fast computation.
   * When the system has barely no penalty for backpropagation, the accuracy of test/tune fluctuates. When increasing penalty (lamdha), the accuracy behaves more stable and trends down when epoch goes on, and shows an available point for early stopping, at around 20 epochs. However, in exchange for stability, the system sacrifices its accuracy by an average of 3% for both tuning and test set.



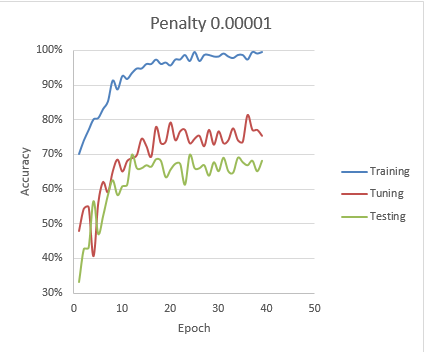
Fig 7: larger penalty

Fig 8: small penalty