Task Sheet 2, Solution

Cristiano Alessandro (alessandro@ifi.uzh.ch), Mathias Weyland (mathias.weyland@uzh.ch)

Solutions

1. Both the Octave/Matlab as well as the pen and paper solution is provided below:

Octave and Matlab

Nobody noted the scalar factor of 2 with $\beta = 1$, so we give the results for $\beta = \frac{1}{2}$ instead to match your solution.

a)
\[
\text{ativ} = @(W,x)(1 + \exp(-W*x)).^-1;
\]
b)
\[
\text{outPut}=@(Wa,Wb,x)\text{ativ}(Wb,[\text{ativ}(Wa,x);1]);
\]
c)
\[
\text{changeW2}=-\text{delta}.*\text{err}.*\text{deriv}.*(Vj');
\]
d)
\[
\text{changeW1}= ...
-\text{delta}*[\text{sumation}(1)*\text{ones}(1,3);\text{sumation}(2)*\text{ones}(1,3)].*[\text{ini'};\text{ini'}].*\text{deriv2};
\]
e) The smaller the learning rate, the slower the convergence.

Pen and paper

Note that we are assuming $\beta = 2$.

a)
\[
\begin{align*}
  h_1(i_1, i_2) &= g(i_1w_{11} + i_2w_{12} + w_{13}) = \frac{1}{1 + \exp(-2(i_1w_{11} + i_2w_{12} + w_{13}))} \\
  h_2(i_1, i_2) &= g(i_1w_{21} + i_2w_{22} + w_{23}) = \frac{1}{1 + \exp(-2(i_1w_{21} + i_2w_{22} + w_{23}))}
\end{align*}
\]
b) 
\[ o(i_1, i_2) = g(h_1(i_1, i_2)w_o1 + g_2(i_1, i_2)w_o2 + w_o3) = \frac{1}{1+\exp(-2(h_1(i_1, i_2)w_o1 + g_2(i_1, i_2)w_o2 + w_o3))} \]

c) Assuming \( o = o(i_1, i_2) \) and \( h_j = h_j(i_1, i_2) \):
\[ \Delta w_{oj} = 2\eta o(1 - o)(\zeta - o)h_j \]

d) 
\[ \Delta w_{jk} = 4\eta h_j(1 - h_j)w_{oj} \cdot o(1 - o)(\zeta - o)i_k \]
e) Table for \( \beta = 2 \):

| \( w_{11} \) | \( w_{12} \) | \( w_{13} \) | \( w_{21} \) | \( w_{22} \) | \( w_{23} \) | \( w_{o1} \) | \( w_{o2} \) | \( w_{o3} \) | \( i_1 \) | \( i_2 \) | \( h_1 \) | \( h_2 \) | \( o \) | \( \Delta w_{11} \) | \( \Delta w_{12} \) | \( \Delta w_{13} \) | \( \Delta w_{21} \) | \( \Delta w_{22} \) | \( \Delta w_{23} \) | \( \Delta w_{o1} \) | \( \Delta w_{o2} \) | \( \Delta w_{o3} \) |
| 0.02 | -0.05 | 0.003 | -0.03 | 0.03 | -0.05 | -0.01 | 0.01 | -0.03 | 0 | 1 | 0.4502 | 0.5135 | 0.4853 | 0 | -0.6364 \times 10^{-3} | 0 | 0.6423 \times 10^{-3} | -0.6364 \times 10^{-3} | 0.6423 \times 10^{-3} | 0.0579 | 0.0660 | 0.1286 |

2. Learning:
a) The trained neural network fits features of the training data that are not related to the model to be learned (e.g. noise). As a result, the trained model will generalize badly.

b) The training data should not be used to assess the quality of the learned model. A new dataset should be used to check if the learning procedure has captured the model underlying the data (as opposed to the specific features of the training samples only).

c) Please see p. 42 of the script.

d) N-fold cross validation can be used to estimate the generalization error. The generalization-vs-#hidden-unit plot can be computed by sequentially adding a hidden unit and evaluating the generalization error. The optimal number of unit correspond to the minimum of such a plot.

3. The momentum can help to speed up learning in flat areas of the error surface. Furthermore it can help to find the way out of (flat) local minima. By the same mechanisms, though, it can also cause the learning algorithm to overshoot out of the global minimum.

4. It becomes 0 in both cases. As a consequence, the deltas also become zero and no error will be propagated (although the error itself could be very large). To avoid this add a small constant to the derivative. This way the deltas are only zero if the error is zero.

5. Cascade correlation:

   a) Please refer to p. 43–45 of the script.

   b) It builds the network in stage: hidden nodes are added only when it is needed based on the data to be approximated. Therefore it ameliorates the problem of overfitting by automatically “growing” the optimal network architecture.

6. More cascade correlation:

   a) With a higher patience parameter, we get a lower final error, higher number of epochs used, and less hidden nodes are recruited. Lower patience parameter, we get a higher final error, lower number of epochs used, and more hidden nodes are recruited.

   b) In general, cascade correlation algorithm is faster than back-propagation in reducing the error.