ECE/CS 559 - Fall 2017 - Some backpropagation practice.

Erdem Koyuncu

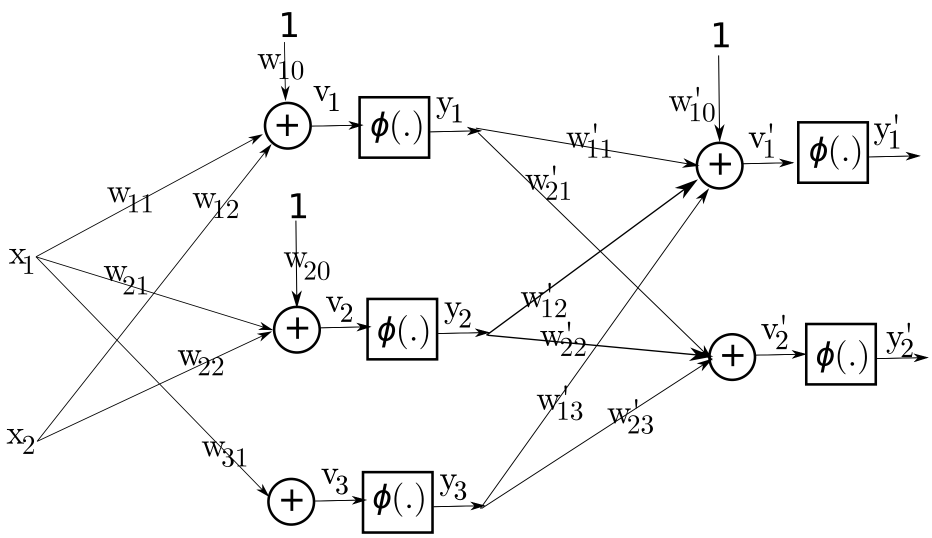
1. Consider the network below. Inputs x_1, x_2 result in the outputs y'_1 and y'_2. Let

$$E = \frac{1}{4}\left[(d_1 - y'_1)^4 + 2(d_2 - y'_2)^2\right] + w_{22}^2,$$

where d_1 and d_2 are constants. Note that E is different compared to the usual squared-error distortion function. Using the backpropagation algorithm, write down the expressions for

$$\frac{\partial E}{\partial w'_{10}}, \frac{\partial E}{\partial w_{22}}, \text{ and } \frac{\partial E}{\partial w'_{13}}$$

as functions of ϕ, the weights, and d_1, d_2.



2. Consider a single neuron with m inputs and a single bias term, i.e. if y is the output of the neuron, and $x_1, \ldots, x_m$ are the inputs, we have $y = \phi(\sum_{i=1}^{m} w_i x_i + w_0)$ for weights $w_0, \ldots, w_m$. Suppose that ϕ is the sigmoid activation function, i.e. $\phi(x) = \beta/(1 + e^{-\alpha x})$, where $\alpha, \beta > 0$. For d being the desired output for some data sample $\mathbf{x} = [x_1 \cdots x_m]^T$, we define the cross-entropy error as

$$C = -d \log y - (1 - d)\log(1 - y).$$

Calculate $\nabla C = [\frac{\partial C}{\partial w_0} \cdots \frac{\partial C}{\partial w_m}]$. Make sure you derive all derivatives.

3. True or False? Briefly justify your answer.
   - The backpropagation algorithm (with gradient descent) can always achieve the optimal solution (the weights that minimize the cost function) with online learning.
   - The backpropagation algorithm can always achieve the optimal solution (the weights that minimize the cost function) with offline learning.
   - The backpropagation algorithm always converges with online learning.

- The backpropagation algorithm always converges with offline learning.
- Consider an arbitrary neural network with a single output. The desired output is 0 for patterns of class C_0, and 1 for patterns of class C_1. The training patterns are chosen from $C_0 \cup C_1$. The backpropagation algorithm always converges when the input patterns are linearly separable, regardless of the type or learning (online or offline).
- The backpropagation algorithm cannot be used for function interpolation.
- The average error achieved by backpropagation with regularization is always better than backpropagation without regularization.
- Gradient descent/backpropagation is used to train multilayer networks as Newton's method is only applicable to single-layer networks.