

# Neural Network Classification

Stephen Vardeman  
Analytics Iowa LLC  
ISU Statistics and IMSE

## Conditional on $\mathbf{x}$ class probabilities

In  $K$ -class classification problems, it is typical to use  $K$  output nodes and for  $\mathbf{w} = (w_1, w_2, \dots, w_K)$  the vector of linear combinations of outputs from the final hidden layer, compute the outputs not simply using a single entry of  $\mathbf{w}$  for each, but rather using all entries. That is, it is typical to set  $K$  outputs to be

$$g_k(\mathbf{w}) = \frac{\exp(w_k)}{\sum_{l=1}^K \exp(w_l)}$$

This vector function of (vector)  $\mathbf{w}$  is usually referred to as the "softmax" function, and produces a probability vector as output. Its entries serve as estimates of class probabilities for the given vector of inputs.

## Neural net *0-1* loss classifier

The 0-1 loss classifier corresponding to this set of estimated class probabilities is then

$$\hat{f}(\mathbf{x}) = \arg \max_k g_k$$

(where it is understood that the  $k$ th probability,  $g_k$ , depends upon the input  $\mathbf{x}$  through the neural net compositions of functions and the final use of the softmax function).