Two points on ANN : Number of neurons and Input Scaling

September 23, 2013

Input Scaling

Consider the sigmoid function

$$f(x) = \frac{1}{1+e^{-x}}$$

Take the derivative

$$\frac{df(x)}{dx} = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= \frac{1}{(1+e^{-x})} \cdot \frac{e^{-x}}{(1+e^{-x})}$$

$$= f(x)\frac{1+e^{-x}-1}{(1+e^{-x})}$$

$$= f(x)\left[1-f(x)\right]$$

Then when $x = 5$ and 6

$$f(5) = 0.993 \; f(6) = 0.997$$

Then

$$\frac{df(x)}{dx}|_{x=5} = 0.006 \qquad \frac{df(x)}{dx}|_{x=6} = 0.0029$$

The rate of change of function around $x = 5$, 6 is so small that there is nearly no difference.

Thus, for processing element with sigmoid function as the activation function, the input of that processing unit must be scaled, otherwise that processing element can not make differentiation on inputs with large values.

Number of neurons in the hidden layers

How to determine how many hidden neurons in hidden layers ? There is no general rule. Normally 1 to 2 hidden layers with few neurons is good enough.

More neuron not neccessarily means better the ANN is. If less number of neurons can achieve the training, using less neurons is better since there is lower computation burden, the fewer neurons, the fewer weights for update.

$-END-$