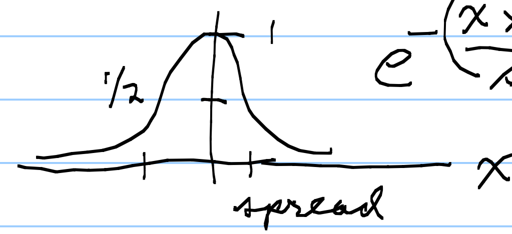
EE 434

Note Title 2/19/2004

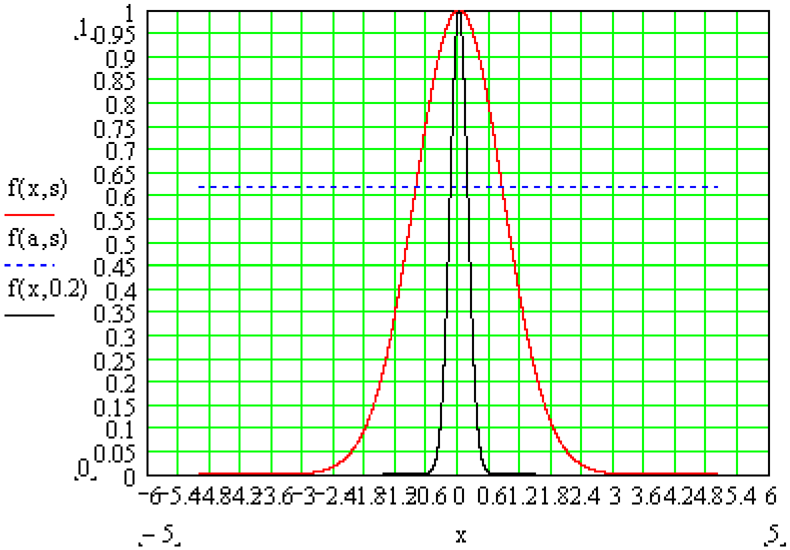
For radial basis function (RBF) networks use newrb

[net,tr] = newrb(P,T,goal,spread,MN,DF)

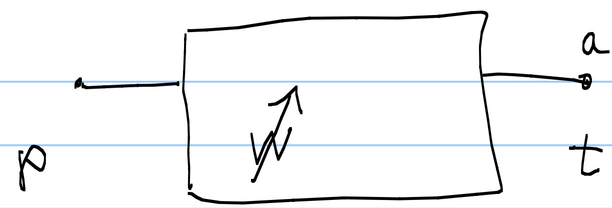



$$e^{-\left(\frac{x \times 0.8326}{\text{spread}}\right)^2} = \text{RBF}$$

first layer biases are all set to 0.8326/spread, resulting in radial basis functions that cross 0.5 at weighted inputs of +/- spread



f(x,s)
f(a,s)
f(x,0.2)
x




know these as given $p$, $t$ an exemplars pairs for training

fix the weights

insert the vector $p$, compare the result $a$ with $t$, if differs then update $W$

define error by

transpose

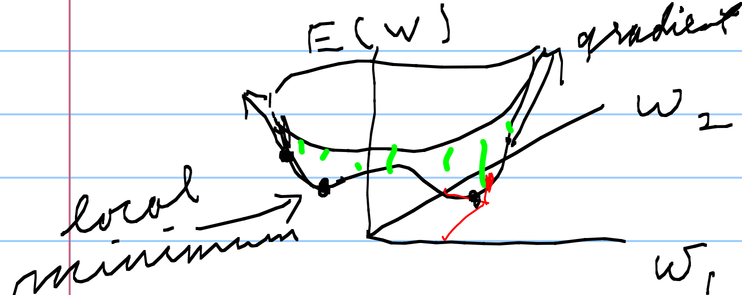
$$E(W) = (t-a)^T(t-a)$$ = sum of squares of component differences

$$= (t-a(W))^T(t-a(W))$$

Ex: if $t = \begin{bmatrix} 2.1 \\ -3.3 \end{bmatrix}$ and $a = \begin{bmatrix} 1.1 \\ 4.1 \end{bmatrix}$ ; $t - a = \begin{bmatrix} 1.0 \\ -7.4 \end{bmatrix}$

$(t-a)^T = [1.0, -7.4]$

$\therefore \; E = (t-a)^T (t-a) = [1.0, -7.4] \begin{bmatrix} 1.0 \\ -7.4 \end{bmatrix} = (1.0)^2 + (-7.4)^2$




$w = [w_1, w_2)$

desire to minimize $E(w)$ by changing weights here have put all weights into a vector

$$E(w^{new}) = E(w^{old}) + \nabla E\Big|_{w^{old}}^{T} (w^{new} - w^{old}) + \text{higher order term}$$

$\therefore$ choose $w^{new} - w^{old}$ be the negative of the gradient $\nabla E \times \alpha$ ; $0 < \alpha < 1$

$$E(w^{new}) = E(w^{old}) - \nabla E^{T} \cdot \nabla E \cdot \alpha + \cdots$$

where $\nabla E = \begin{bmatrix} \partial E / w_1 \\ \partial E / w_2 \\ \vdots \\ \partial E / w_{\text{last weight}} \end{bmatrix}$

$\therefore$ choose

$$W^{new} = W^{old} - \alpha \nabla E \Big|_{W = W^{old}}$$

this the weight update "formula" given a scalar error function $E(W)$.

output layer of a multilayer feed forward network

$$\cdots\cdots \; a^{M-1} \xrightarrow{\;W^M\;} n^M \xrightarrow{\;f^M(\cdot)\;} a = a^M$$

$$a^M = f^M(W^M a^{M-1})$$

$$W^{M\,new} = W^{M\,old} - \alpha \nabla\Big|_{W^M} (t - a^M)$$

$j$th component of $a$ & $t$

$\nabla_{w^M}\left|(t - a^M(w^M))_j\right.$ ; $i$th component, $w_i^M = w_i$

$$\frac{\partial}{\partial w_i}(t_j - a_j(w^M)) = -\frac{\partial}{\partial w_i} a_j(w^M)$$

$$\frac{\partial a_j(w^M)}{\partial w_i^M} = \frac{\partial f_j^M(w^M a^{M-1})}{\partial w_i^M} = \left.\frac{d f_j^M(n_j)}{d n_j}\right|_{n_j^{old}} \cdot \frac{\partial n_j}{\partial w_i^M}$$

here $\displaystyle\frac{\partial n_j}{\partial w_i^M} = \frac{\partial \sum_i w_{ji} a_i^{M-1}}{\partial w_i^M} = a_i^{M-1}$ for $w_{ji} = w_i^M$

here $\dfrac{df}{dn}$ is known once activation functions are known