Neil E Cotter
13 Apr 1994

tool: The following gaussian radial basis function network satisfies the Stone–Weierstrass theorem and is, therefore, capable of universal approximation, (on compact domain, D).

$$\mathcal{F} = \Big\{ f(\vec{x}) = \sum_{j=1}^{M} w_j \, r_j(\vec{x}) \Big\}, \qquad r_j(\vec{x}) = e^{-a_j |\vec{x} - \vec{x}_{0j}|^2}$$

where $w_j \in \mathbb{R}$ (real #'s)
$a_j \in \mathbb{Z}^+$ (integers ≥ 0)
$x_{0j} \in \mathbb{Q}$ (rational numbers) $\}$

pf: To satisfy Stone-Weierstrass we must ~~have~~ satisfy the following requirements:

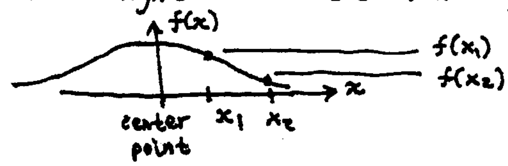
- *Domain D is compact space of N dimensions.*
  (We specified that D is compact and $\vec{x}$ has N dimensions. So this requirement is satisfied.)

- *$\mathcal{F}$ is a set of continuous real-valued functions on D.*
  ($f(\vec{x}) \in \mathcal{F}$ is always a sum of gaussians which are continuous and real-valued on D. Thus, $f(\vec{x})$ is continuous and real-valued on D. So this requirement is satisfied.)

- *The constant function $f(\vec{x}) = 1$ is in $\mathcal{F}$.*
  (To create this $f(\vec{x})$ let $M = 1$ and
  $w_1 \cancel{a_1} = 1$
  $a_1 = 0$
  $x_{01} = \vec{0}$.
  Then $f(\vec{x}) = 1 \cdot e^{-0|\vec{x} - \vec{0}|^2} = e^0 = 1$
  for all $\vec{x}$. So $f(\vec{x}) = 1$ is in $\mathcal{F}$.)

M. E Cotter.
13 Apr 1994.

- For any two points $\vec{x}_1 \neq \vec{x}_2$ in D, there is an f in $\mathcal{F}$ such that $f(\vec{x}_1) \neq f(\vec{x}_2)$.
  (Given $\vec{x}_1$ and $\vec{x}_2$ we create the required $f(\vec{x})$ by choosing a center point such that $\vec{x}_1$ and $\vec{x}_2$ lie at different distances from the center point. For example, in one dimension we might have the following picture:



  Clearly, we can always find such a center point. We let $f(\vec{x})$ consist of this single gaussian bump. So the requirement is satisfied.)

- If f and g are any two functions in $\mathcal{F}$ then $af + bg$ is in $\mathcal{F}$ for any two real numbers a and b.
  (We show this directly by summing f and g and showing that the result can be written as a third gaussian radial basis function net in $\mathcal{F}$.

Given $$f \equiv f(\vec{x}) = \sum_{j=1}^{M_1} w_j\, r_j(\vec{x})$$

$$g \equiv g(\vec{x}) = \sum_{k=1}^{M_2} w_k\, r_k(\vec{x})$$

$$af + bg = \sum_{j=1}^{M_1} a w_j\, r_j(\vec{x}) + \sum_{k=1}^{M_2} b w_k\, r_k(\vec{x})$$

$$'' = \sum_{\ell=1}^{M_1+M_2} w_\ell\, r_\ell(\vec{x})$$

where $w_\ell = a w_{j=\ell}$ and $r_\ell = r_{j=\ell}$ for $\ell \leq M_1$

$w_\ell = b W_{k=\ell-M_1}$ and $r_\ell = r_{k=\ell-M_1}$ for $M_1 < \ell$

N. E. Cotter
13 Apr 1994

We see that $af + bg = \sum_{\ell=1}^{M_1+M_2} w_\ell r_\ell(\vec{x})$ is a gaussian radial basis function network in $\mathcal{F}$. Thus, the addition and scaling requirement is satisfied.)

note: When we said $w_\ell = a w_{j=\ell}$ we used the condition that w_ℓ can be any real number. If we had said w values are integers we could have $a w_j \neq$ integer since "a" may be any real number. In that case w_ℓ would not be an integer and our new network would not have integer weights like f and g. Thus, our $af+bg$ would not be in $\mathcal{F}$. This is why we allow w's to be any real numbers.

- If f and g are any two functions in $\mathcal{F}$ then fg is in $\mathcal{F}$.

(Again, we show this by directly computing fg and showing the result is a function in $\mathcal{F}$. Let f and g be defined as above. Then we have

$$fg = \sum_{j=1}^{M_1} w_j r_j(\vec{x}) \sum_{k=1}^{M_2} w_k r_k(\vec{x})$$
$$= \sum_{j=1}^{M_1} \sum_{k=1}^{M_2} w_j w_k \, r_j(\vec{x}) r_k(\vec{x})$$

Consider $r_j(\vec{x}) r_k(\vec{x})$ in detail. To simplify the discussion and notation, assume $\vec{x} = x$ is one-dimensional.

$$r_j(x) r_k(x) = e^{-a_j (x - x_{oj})^2} \cdot e^{-a_k (x - x_{ok})^2}$$
$$= e^{-[a_j (x - x_{oj})^2 + a_k (x - x_{ok})^2]}$$

© E Cotter
13 Apr 1994.

Consider just the exponent:

$$a_j (x - x_{oj})^2 + a_k (x - x_{ok})^2 =$$

$$(a_j + a_k) x^2 - 2(a_j x\, x_{oj} + a_k x\, x_{ok}) + a_j x_{oj}^2 + a_k x_{ok}^2$$

We can rewrite this as follows:

$$(a_j + a_k)(x - x_{ojk})^2 + c$$

where $x_{ojk} = \dfrac{a_j x_{oj} + a_k x_{ok}}{a_j + a_k}$ $\in \mathbb{Q}$ (rationals) so center point is still rational

$$c = a_j x_{oj}^2 + a_k x_{ok}^2 - \frac{(a_j x_{oj} + a_k x_{ok})^2}{a_j + a_k}$$

So we have

$$fg = \sum_{j=1}^{M_1} \sum_{k=1}^{M_2} w_j W_k \, e^{-(a_j + a_k)(x - x_{ojk})^2} e^{-c}$$

Define $\omega_{jk} \equiv w_j W_k e^{-c}$ and $\alpha_{jk} \equiv a_j + a_k$.

Note that $\omega_{jk} \in \mathbb{R}$ (real #'s) and $\alpha_{jk} \in \mathbb{Z}$ (ints) so we will get a function in $\tilde{\mathcal{G}}$ later on.

$$fg = \sum_{j=1}^{M_1} \sum_{k=1}^{M_2} \omega_{jk} \, e^{-\alpha_{jk} (x - x_{ojk})^2}$$

We can rewrite this double Σ as a *single* Σ. We *implicitly define* ω_ℓ *to correspond to* ω_{jk} etc. *in the obvious way, i.e. write out the double sum and number the terms with* $\ell = 1, \ldots, M_1 M_2$.

$$fg = \sum_{\ell=1}^{M_1 M_2} \omega_\ell \, e^{-\alpha_\ell (x - x_{o\ell})^2}$$

Mi[illegible] Cotter
13 Apr 1994

We see that f_g is in $\mathcal{F}$ and our proof is complete.)

comment: This proof shows that we can approximate arbitrary functions with arbitrary accuracy, but it doesn't say how many neurons we will need. Thus, we cannot specify the network in advance of knowing what function we want to approximate and still be sure of our final accuracy.

ref: E. J. Hartman, J. D. Keeler, J. M. Kowalski
Layered Neural Networks with Gaussian Hidden Units as Universal Approximations, Neural Computation 2, pp 210-215 (1990)