

Mean Dimension of Radial Basis Functions

Art B. Owen*

Stanford University

Based on joint work with:

Christopher Hoyt, Stanford University

Opinions are my own, and not those of Stanford, the NSF, or Hitachi, Ltd.

Radial basis functions

$$\phi(\|\mathbf{x} - \mathbf{c}\|) \quad \text{for } \mathbf{x}, \mathbf{c} \in \mathbb{R}^d$$

E.g., scattered data interpolation

$$f(\mathbf{x}) \approx \tilde{f}(\mathbf{x}) = \sum_{i=1}^n \beta_i \phi(\|\mathbf{x} - \mathbf{x}_i\|)$$

Versus ridge functions

$$\phi(\mathbf{x}^\top \boldsymbol{\theta}) \quad \boldsymbol{\theta}^\top \boldsymbol{\theta} = 1$$

What we find

Some RBFs become essentially additive in high dimensions

Why we care

That limits their usefulness as

approximators

covariance functions

Other RBFs

Not necessarily additive

Same for ridge functions

Concentration of measure

High dimensional and Lipschitz

⇒ Nearly constant

Donoho (2000)

Variation around constant

Depends

RBF interpolation

$$\tilde{f}(\mathbf{x}) = \sum_{i=1}^n \beta_i \phi(\|\mathbf{x} - \mathbf{x}_i\|)$$

Scattered data \mathbf{x}_i

e.g., IID

Choose $\beta = (\beta_1, \dots, \beta_n) \in \mathbb{R}^n$ so that

$$\tilde{f}(\mathbf{x}_i) = f(\mathbf{x}_i)$$

Doable for

Strictly positive definite RBFs

some 'conditionally positive definite' RBFs

See [Fasshauer \(2007\)](#)

Polynomial interpolation not always effective

Some strictly PSD RBFs

$r = \ \mathbf{x} - \mathbf{c}\ $	e.g., $\mathbf{c} = \mathbf{x}_i$
$\phi(r) = \exp(-\theta^2 r^2)$	Gaussian / squared exponential
$\phi(r) = (1 + \theta^2 r^2)^p$ $p < 0$	Generalized inverse multiquadric
$\phi(r) = (1 + \theta^2 r^2)^{-1/2}$	Inverse multiquadric
$\phi(r) = (1 + \theta^2 r^2)^{1/2}$	Multiquadric Hardy (1971)

Schoenberg condition

$$\phi(r) = \int_0^\infty e^{-r^2 t^2} \mu(dt)$$

For strictly PSD RBF, any $d \geq 1$

Fasshauer (2007) Theorem 3.8

$\phi(\cdot) \geq 0$ nondecreasing

ANOVA

For $x_j \sim F_j$ independent & $0 < \mathbb{E}(f(\mathbf{x})^2) < \infty$

Fisher & Mackenzie, Hoeffding, Sobol', Efron & Stein

Notation

$[d] \equiv \{1, 2, \dots, d\}$ For $u \subseteq [d]$, $\mathbf{x}_u = (x_j)_{j \in u}$

ANOVA decomp

$$f(\mathbf{x}) = \sum_{u \subseteq [d]} f_u(\mathbf{x})$$

$f_u(\mathbf{x})$ only depends on \mathbf{x}_u

Recursive definition

$$f_{\emptyset}(\mathbf{x}) = \mathbb{E}(f(\mathbf{x}))$$

$$f_u(\mathbf{x}) = \mathbb{E}\left(f(\mathbf{x}) - \sum_{v \subsetneq u} f_v(\mathbf{x}) \mid \mathbf{x}_u\right) = \mathbb{E}(f(\mathbf{x}) \mid \mathbf{x}_u) - \sum_{v \subsetneq u} f_v(\mathbf{x})$$

ANOVA continued

Recall $f(\mathbf{x}) = \sum_{u \subseteq [d]} f_u(\mathbf{x})$

From orthogonality

$$\sigma^2 = \text{Var}(f(\mathbf{x}))$$

$$\sigma_u^2 = \text{Var}(f_u(\mathbf{x}))$$

$$\sigma^2 = \sum_{u \subseteq [d]} \sigma_u^2$$

Mean dimension

$$\nu(f) = \sum_{u \subseteq [d]} |u| \frac{\sigma_u^2}{\sigma^2}$$

E.g., for $d = 3$

$$\nu(f) = \frac{\sigma_{\{1\}}^2 + \sigma_{\{2\}}^2 + \sigma_{\{3\}}^2 + 2\sigma_{\{1,2\}}^2 + 2\sigma_{\{1,3\}}^2 + 2\sigma_{\{2,3\}}^2 + 3\sigma_{\{1,2,3\}}^2}{\sigma^2}$$

Additivity

$$R_{\text{add}}^2 = \frac{1}{\sigma^2} \sum_{j=1}^d \sigma_{\{j\}}^2$$

$$\nu(f) \leq 1 + \epsilon \implies R_{\text{add}}^2 \geq 1 - \epsilon$$

Sobol' indices

$$\bar{\tau}_j^2 = \sum_{v: j \in v} \sigma_v^2$$

Jansen

$$\bar{\tau}_j^2 = \frac{1}{2} \mathbb{E}((f(\mathbf{x}) - f(\mathbf{z}_j: \mathbf{x}_{-j}))^2)$$

Liu & O

$$\underbrace{\sum_{u \subseteq [d]} |u| \sigma_u^2}_{2^d \text{ terms}} = \sum_{j=1}^d \underbrace{\bar{\tau}_j^2}_{d \text{ terms}}$$

Pick freeze

For $\mathbf{z}_j: \mathbf{x}_{-j}$ replace x_j by independent z_j

Main result

For generalized multiquadrics, independent x_j

$$f(\mathbf{x}) = (\textcolor{red}{a} + \|\mathbf{x} - \mathbf{c}\|^2)^p \quad a \geq 0 \quad p < 1$$

Then under conditions

$$\nu(f) = 1 + O(\textcolor{red}{d}^{-1}) \quad \text{With explicit constant}$$

More notation

$$z_j = \begin{cases} a + (x_1 - c_1)^2, & j = 1 \\ (x_j - c_j)^2, & j > 1. \end{cases}$$

Study

Mean dim of

$$\left(\sum_{j=1}^d z_j \right)^p$$

equals mean dim in \mathbf{x}

Transforming to z

$$z_j = \begin{cases} a + (x_1 - c_1)^2, & j = 1 \\ (x_j - c_j)^2, & j > 1. \end{cases}$$

Define

$$\begin{aligned} \mu_j &= \mathbb{E}(z_j) & \sigma_j^2 &= \text{Var}(z_j) & \mu_j^{(k)} &= \mathbb{E}((z_j - \mu_j)^k) \\ \mu_{1:d} &= \sum_{j=1}^d \mu_j & \sigma_{1:d}^2 &= \sum_{j=1}^d \sigma_j^2 & \mu_{1:d}^{(k)} &= \sum_{j=1}^d \mu_j^{(k)} \end{aligned}$$

Under conditions (up to 6 moments)

$$\nu(f) \leq 1 + \frac{(p-1)^2}{2} \frac{\sigma_{1:d}^2}{(\mu_{1:d})^2} + O\left(\frac{1}{d^2}\right)$$

Weaker conditions

$$\nu(f) \rightarrow 1$$

Proof strategy

$$\nu(f) = \frac{\sum_{j=1}^d \bar{\tau}_j^2}{\sigma^2}$$

$$\limsup_{d \rightarrow \infty} \frac{\sum_{j=1}^d \bar{\tau}_j^2}{p^2 \sigma_{1:d}^2 / (\mu_{1:d})^2} \leq 1$$

$$\liminf_{d \rightarrow \infty} \frac{\text{Var}(f(\mathbf{z}))}{p^2 \sigma_{1:d}^2 / (\mu_{1:d})^2} \geq 1$$

Ingredients

Taylor expansions

moment bounds

Jansen identity

negative moment bounds

central limit theorem

Gaussian RBF

$$f(\mathbf{x}) = \prod_{j=1}^d \exp(-(x_j - c_j)^2 / \vartheta^2)$$

Mean dim of a product

For $f(\mathbf{x}) = \prod_{j=1}^d g_j(x_j)$

$$\nu(f) = \frac{\sum_{j=1}^d \rho_j}{1 - \prod_{j=1}^d (1 - \rho_j)}$$
$$\rho_j = \text{Var}(z_j) / \mathbb{E}(z_j^2) \in [0, 1]$$

O (2003)

Then

$$\frac{\partial}{\partial \rho_k} \nu(f) \geq 0$$

Hoyt & O (2023)

So \dots larger $\rho_j \Rightarrow$ larger ν

Gaussian RBF theorem

Pick θ to get any $0 < \rho_j < 1$

\Rightarrow get any $1 < \nu(f) < d$

Independent x_j with

$$z_j = \frac{(x_j - c_j)^2}{\theta^2} \quad \text{Var}(x_j) > 0 \quad \text{PDFs } h_j(x_j) \leq M$$

Then

$$f(\mathbf{x}) = f(\mathbf{z}) = \prod_{j=1}^d \exp(-z_j)$$

can attain any $1 < \nu(f) < d$ by choosing θ

Speculation

This may be why Gaussian RBFs dominate other RBFs in machine learning

Ridge in RQMC/QMC

$$\boldsymbol{x} \sim \mathcal{N}(0, I_d) \quad f_d(\boldsymbol{x}) = g(\theta^\top \boldsymbol{x})$$

Family of test functions with same mean and variance

In any dimension

For Lipschitz $g(\cdot)$

$$\sup_{d \geq 1} \nu(f_d) < \infty$$

Get modest limiting mean dimension, e.g., ≤ 2 or 3.

Speculation

This may be why ridge functions dominate Gaussian RBFs in machine learning

Covariance functions

Gaussian process (GP) model for $f(\mathbf{x})$ on $\mathbf{x} \in \mathbb{R}^d$

$$\mathbb{E}(f(\mathbf{x})) = \mu(x) \quad \text{Cov}(f(\mathbf{x}), f(\tilde{\mathbf{x}})) = \Sigma(\mathbf{x}, \tilde{\mathbf{x}})$$

For a GP

$$\Sigma(\mathbf{x}, \tilde{\mathbf{x}}) = \phi(\|\mathbf{x} - \tilde{\mathbf{x}}\|)$$

ϕ is additive $\Rightarrow f$ is additive

Additive f is a sum of independent GPs

Speculation

This may be why Gaussian covariances dominate multiquadrics in GP models

Keister function

$$\int_{\mathbb{R}^d} e^{-\|\mathbf{x}\|^2} \cos(\|\mathbf{x}\|) d\mathbf{x}$$

Keister (1996), Capstick & Keister (1996)

Resembles problems from physics Rewrite

$$\mathbb{E}\left(\cos\left(\frac{\|\mathbf{x}\|}{2}\right)\right) \quad \mathbf{x} \sim \mathcal{N}(0, I_d)$$

QMC context

This function treats all inputs symmetrically

Yet we still see good QMC integration

Papageorgiou & Traub (1997)

Mean dimension

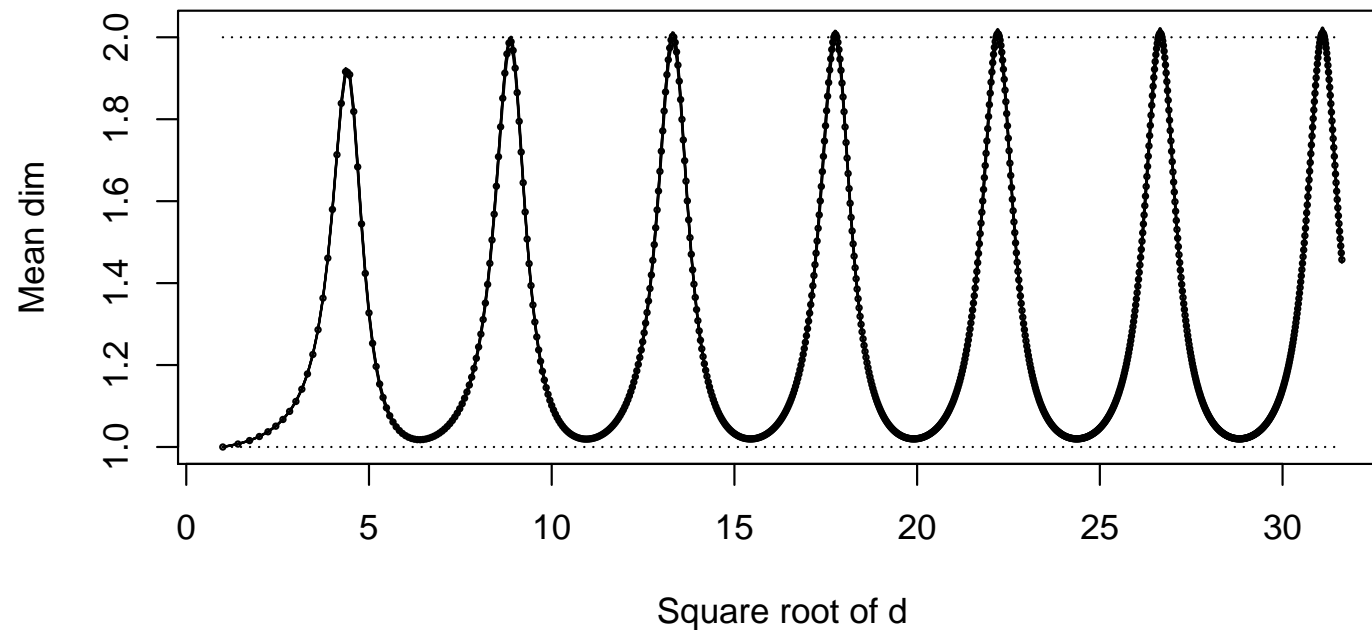
It has modest mean dimension in any d

and smoothness

⇒ QMC accuracy not surprising

Keister mean dimension

Keister's function: mean dimension vs nominal



$\nu(f)$ vs \sqrt{d}

Five independent QMC estimates overlaid

Based on **Jansen** identity and 3 χ^2 random variables

Why the oscillation?

$$\|\mathbf{x}\|^2 \sim \chi_{(d)}^2 \approx \mathcal{N}(d, 2d)$$

By delta method (Taylor's theorem)

$$\frac{\|\mathbf{x}\|}{2} \approx \mathcal{N}\left(\frac{\sqrt{d}}{2}, \frac{1}{4}\right)$$

$\approx 99.9\%$ of $\mathcal{N}(\alpha, 1/4)$ is in $\alpha \pm 1.65$

Distn of $\|\mathbf{x}\|/2$ uses \approx half period of cosine

Centering

$$\frac{\sqrt{d}}{2\pi} \approx \text{integer} \quad \Rightarrow \quad \cos(\cdot) \text{ nearly pure quadratic} \quad \Rightarrow \quad \nu(f) \approx 2$$

$$\frac{\sqrt{d}}{2\pi} \approx \text{integer} + \frac{1}{2} \quad \Rightarrow \quad \cos(\cdot) \text{ nearly linear} \quad \Rightarrow \quad \nu(f) \approx 1$$

Thanks

- Christopher Hoyt, co-author
- Masayoshi Mase and Naofumi Hama, comments
- NSF IIS-1837931 and DMS-2152780
- Hitachi, Ltd.
- Session organizers: Mathias Sonnleitner & Mario Ullrich
- MCM organizers