

2

Núcleos: suas propriedades e classificações

O objetivo desse capítulo é o de apresentar as funções núcleos (7), suas propriedades (10) e suas classificações (3).

2.1

Núcleos no espaço de Hilbert

Um espaço de Hilbert é um espaço vetorial munido de um produto interno e cuja métrica gerada por esse produto interno o torna um espaço completo, ou seja nele todas as seqüências de Cauchy convergem. Para lembrar, uma seqüência de pontos $\{\mathbf{y}_n\}$ em \mathcal{F} é de Cauchy se para todo $\epsilon > 0$, existe um índice n_0 tal que para quaisquer índices m e n maiores que n_0 , tem-se que a distância entre \mathbf{y}_m e \mathbf{y}_n é menor do que ϵ .

Os núcleos $k : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ são funções que correspondem ao produto interno no espaço de Hilbert \mathcal{F} da imagem dos pontos \mathbf{x} e \mathbf{z} por uma função $\phi : \mathbb{R}^n \rightarrow \mathcal{F}$, ou seja

$$k(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle .$$

Na prática, as funções núcleos representam implicitamente o mapeamento feito pela função ϕ , ou seja, se define qual núcleo usar e não qual ϕ usar. Um ponto interessante agora é saber como construir funções que são núcleos.

É importante observar que pela simetria do produto interno, um núcleo deve ser simétrico, isto é:

$$k(\mathbf{x}, \mathbf{z}) = k(\mathbf{z}, \mathbf{x}).$$

Além disso, ele deve satisfazer a desigualdade de Cauchy-Schwartz:

$$k^2(\mathbf{x}, \mathbf{z}) \leq k(\mathbf{x}, \mathbf{x})k(\mathbf{z}, \mathbf{z}).$$

Construir uma função k que satisfaça essas duas propriedades não é tão difícil, mas isso não garante a existência do espaço de Hilbert \mathcal{F} . Porém, Mercer em 1909 (7) provou que uma condição necessária e suficiente para que uma função simétrica seja um núcleo é que ela seja positiva definida. Isso significa que para qualquer conjunto $\mathbf{x}_1, \dots, \mathbf{x}_N$ de pontos e para quaisquer números reais

$\lambda_1, \dots, \lambda_N$ essa função deve satisfazer:

$$\sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j k(\mathbf{x}_i, \mathbf{x}_j) \geq 0.$$

Os núcleos, que são na realidade funções simétricas definidas positivas, são bastante conhecidos por terem várias propriedades algébricas interessantes. E o sucesso deles em estatística vem do fato que eles podem ser pensados como funções de covariância.

A seguir serão apresentadas formas de se criar núcleos a partir de outros núcleos.

Primeiramente, se k_1 e k_2 são dois núcleos e α_1 e α_2 são dois números reais positivos, então

$$k(\mathbf{x}, \mathbf{z}) = \alpha_1 k_1(\mathbf{x}, \mathbf{z}) + \alpha_2 k_2(\mathbf{x}, \mathbf{z})$$

é um núcleo. Além disso,

$$k(\mathbf{x}, \mathbf{z}) = k_1(\mathbf{x}, \mathbf{z}) k_2(\mathbf{x}, \mathbf{z})$$

é também um núcleo. Como consequência dessas duas propriedades, pode-se dizer que qualquer polinômio com coeficientes positivos, $pol^+(x) = \sum_{i=1}^g a_i x^i$ com $x \in \mathbb{R}$, $a_i \in \mathbb{R}^+$ e grau $g \in \mathbb{N}$, avaliados em k_1 também gera um núcleo:

$$k(\mathbf{x}, \mathbf{z}) = pol^+(k_1(\mathbf{x}, \mathbf{z})).$$

Em particular, o núcleo

$$k(\mathbf{x}, \mathbf{z}) = \exp(k_1(\mathbf{x}, \mathbf{z}))$$

pode ser construído considerando a expansão de Taylor da função exponencial em torno da origem.

Se g é uma função real definida em \mathbb{R}^n , então

$$k(\mathbf{x}, \mathbf{z}) = g(\mathbf{x})g(\mathbf{z})$$

é um núcleo.

Se $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^p$ é uma função e $k_3 : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ é um núcleo, então

$$k(\mathbf{x}, \mathbf{z}) = k_3(\psi(\mathbf{x}), \psi(\mathbf{z}))$$

é um núcleo.

E, finalmente, se $A \in \mathbb{R}^{n \times n}$ é uma matriz simétrica definida positiva, então

$$k(\mathbf{x}, \mathbf{z}) = \mathbf{x}^T A \mathbf{z}$$

é um núcleo.

2.2

Classes de núcleos

Genton em (3) agrupou os núcleos em classes, são elas: núcleos estacionários (que por sua vez podem ser isotrópicos ou anisotrópicos e ainda de suporte compacto ou não), núcleos localmente estacionários e núcleos não-estacionários.

2.2.1

Núcleos estacionários

Um *núcleo estacionário* é aquele que invariante por translação, isto é, ele depende da diferença entre os vetores de entrada, e não da posição em que eles estão. Com isso, pode-se dizer que $k : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ é um núcleo estacionário se existe uma função $k_S : \mathbb{R}^n \rightarrow \mathbb{R}$ que, para quaisquer \mathbf{x} e $\mathbf{z} \in \mathbb{R}^n$, vale:

$$k(\mathbf{x}, \mathbf{z}) = k_S(\mathbf{x} - \mathbf{z}).$$

Um núcleo estacionário k_S que depende do sentido e do comprimento do vetor $\mathbf{x} - \mathbf{z}$ é chamado de *anisotrópico*. Nesse trabalho, será dada uma maior ênfase aos núcleos estacionários *isotrópicos*. Nesse caso k_S depende somente da distância entre dois pontos e não da direção, isto é:

$$k(\mathbf{x}, \mathbf{z}) = k_I(\|\mathbf{x} - \mathbf{z}\|).$$

A tabela 2.1 contém exemplos dos mais conhecidos núcleos estacionários que são isotrópicos, todos parametrizados por um número real $\theta > 0$.

Tabela 2.1: Alguns núcleos estacionários e isotrópicos.

Nome do núcleo	$k_I(\ \mathbf{x} - \mathbf{z}\) =$
<i>Circular</i>	$\begin{cases} \frac{2}{\pi} \arccos\left(\frac{\ \mathbf{x} - \mathbf{z}\ }{\theta}\right) - \frac{2}{\pi} \frac{\ \mathbf{x} - \mathbf{z}\ }{\theta} \sqrt{1 - \frac{\ \mathbf{x} - \mathbf{z}\ ^2}{\theta^2}}, & \text{se } \ \mathbf{x} - \mathbf{z}\ < \theta \\ 0, & \text{caso contrário.} \end{cases}$
<i>Racional quadrático</i>	$1 - \frac{\ \mathbf{x} - \mathbf{z}\ ^2}{\ \mathbf{x} - \mathbf{z}\ ^2 + \theta}$
<i>Exponencial</i>	$\exp\left(-\frac{\ \mathbf{x} - \mathbf{z}\ }{\theta}\right)$
<i>Gaussiano</i>	$\exp\left(-\frac{(\ \mathbf{x} - \mathbf{z}\)^2}{\theta}\right)$
<i>Wave</i>	$\begin{cases} \frac{\theta}{\ \mathbf{x} - \mathbf{z}\ } \sin\left(\frac{\ \mathbf{x} - \mathbf{z}\ }{\theta}\right), & \text{se } \ \mathbf{x} - \mathbf{z}\ \neq 0 \\ 1, & \text{caso contrário.} \end{cases}$

Um exemplo de um núcleo estacionário anisotrópico é o *núcleo wavelets* (12). A ideia por trás da teoria de wavelets é de expressar ou aproximar uma função por uma família de dilatações e translações de uma função $h(x)$ que é chamada de wavelet mãe e denotada por $h_{\theta,c}(x)$:

$$h_{\theta,c}(x) = \frac{1}{|\theta|^{\frac{1}{2}}} h\left(\frac{x-c}{\theta}\right)$$

onde $\theta \in \mathbb{R}$ é o coeficiente de dilatação, $c \in \mathbb{R}$ é o coeficiente de translação, e $x \in \mathbb{R}$. Para uma função h ser considerada uma wavelet mãe, ela deve satisfazer a seguinte condição:

$$\int_0^\infty \frac{|H(w)|^2}{|w|} dw < \infty,$$

onde $H(w)$ é a transformada de Fourier de $h(x)$.

Um exemplo concreto dessa função h é dado por:

$$h(x) = \cos(1.75x)e^{-\frac{x^2}{2}}.$$

Quando são considerados pontos de entrada \mathbf{x} em \mathbb{R}^n pode-se utilizar a seguinte função wavelet n -dimensional:

$$h_n(\mathbf{x}) = \prod_{i=1}^n h(x_i),$$

onde x_i é a i -ésima coordenada do vetor $\mathbf{x} \in \mathbb{R}^n$.

Finalmente, o *núcleo wavelet* invariante por translação é definido da seguinte forma:

$$k(\mathbf{x}, \mathbf{z}) = h_n\left(\frac{\mathbf{x} - \mathbf{z}}{\theta}\right)$$

Os *núcleos com suporte compacto* são aqueles que são nulos quando a distância entre os pontos \mathbf{x} e \mathbf{z} é maior do que um certo limiar, como exemplo o núcleo circular contido na tabela 2.1. Essa propriedade é crucial quando o número N de pontos no conjunto \mathcal{S} é grande. Tornar um núcleo sem suporte compacto num núcleo com suporte compacto de uma maneira bruta é uma tarefa arriscada pois o núcleo pode perder a propriedade de ser uma função simétrica definida positiva.

2.2.2

Núcleos localmente estacionários

Um núcleo é *localmente estacionário* quando pode ser escrito na forma:

$$k_L(\mathbf{x}, \mathbf{z}) = k_1\left(\frac{\mathbf{x} + \mathbf{z}}{2}\right) k_2(\mathbf{x} - \mathbf{z}), \quad (2-1)$$

onde k_1 é uma função não negativa e k_2 é um núcleo estacionário. Observe que se k_1 é uma constante positiva k se torna um núcleo estacionário, assim a classe dos núcleos estacionários é um caso particular dessa.

Se for imposta a condição de que $k_2(\mathbf{0}) = 1$, então

$$k_L(\mathbf{x}, \mathbf{x}) = k_1(\mathbf{x})k_2(\mathbf{0}) = k_1(\mathbf{x}).$$

Nesse caso pode-se dizer que k_1 descreve a estrutura global do núcleo. Por outro lado, ao se avaliar

$$k_L(\mathbf{x}/2, -\mathbf{x}/2) = k_1(\mathbf{0})k_2(\mathbf{x}),$$

pode-se dizer que k_2 está fazendo a contribuição local do núcleo.

A partir dessas duas equações acima é possível reescrever a equação (2-1) da seguinte forma:

$$k_L(\mathbf{x}, \mathbf{z}) = k_1\left(\frac{\mathbf{x} + \mathbf{z}}{2}\right) k_2(\mathbf{x} - \mathbf{z}) = \frac{k_L((\mathbf{x} + \mathbf{z})/2, (\mathbf{x} + \mathbf{z})/2)k_L((\mathbf{x} - \mathbf{z})/2, -(\mathbf{x} - \mathbf{z})/2)}{k_L(\mathbf{0}, \mathbf{0})}.$$

Essa nova formulação diz que um núcleo localmente estacionário k é completamente determinado pelos seus valores na diagonal $\mathbf{x} = \mathbf{z}$ e pela antidiagonal $\mathbf{x} = -\mathbf{z}$.

2.2.3

Núcleos não-estacionários

Um núcleo é *não-estacionário* quando depende explicitamente dos dois pontos \mathbf{x} e \mathbf{z} . Um exemplo de núcleo não estacionário é o *núcleo polinomial de grau* $\theta \in \mathbb{N}$:

$$k_{pol}(\mathbf{x}, \mathbf{z}) = (\langle \mathbf{x}, \mathbf{z} \rangle)^\theta.$$

Uma família particular de núcleos não-estacionários é a de núcleos não-estacionários separáveis, na qual o núcleo k pode ser escrito da seguinte forma:

$$k(\mathbf{x}, \mathbf{z}) = k_a(\mathbf{x})k_b(\mathbf{z}),$$

onde k_a e k_b são núcleos estacionários. Essa família de núcleos possui uma propriedade bastante importante em aprendizado estatístico que é: a matriz de Gram $\mathbf{G} \in \mathbb{R}^{N \times N}$ para os pontos de entrada $\{\mathbf{x}_i\}, i = 1 \dots N$, onde $\mathbf{G}_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$, pode ser escrita como um produto tensorial. Essa propriedade é bastante importante para algoritmos de aprendizagem estatística.

2.3

Distâncias com uso de núcleos

Um núcleo $k(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle$ pode ser usado para calcular a distância entre os pontos $\phi(\mathbf{x})$ e $\phi(\mathbf{z})$ no espaço \mathcal{F} , ou seja

$$d_k(\mathbf{x}, \mathbf{z}) = \|\phi(\mathbf{x}) - \phi(\mathbf{z})\| = \sqrt{\langle \phi(\mathbf{x}) - \phi(\mathbf{z}), \phi(\mathbf{x}) - \phi(\mathbf{z}) \rangle}.$$

Para isso basta usar o seguinte desenvolvimento:

$$d_k(\mathbf{x}, \mathbf{z}) = \sqrt{\langle \phi(\mathbf{x}), \phi(\mathbf{x}) \rangle - 2\langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle + \langle \phi(\mathbf{z}), \phi(\mathbf{z}) \rangle},$$

que substituindo a definição de núcleo vira:

$$d_k(\mathbf{x}, \mathbf{z}) = \sqrt{k(\mathbf{x}, \mathbf{x}) - 2k(\mathbf{x}, \mathbf{z}) + k(\mathbf{z}, \mathbf{z})}.$$

Portanto, é possível calcular a distância entre os pontos $\phi(\mathbf{x})$ e $\phi(\mathbf{z})$ sem precisar especificar a ϕ mas sim o núcleo k .

Muitas vezes, em núcleos estacionários isotrópicos tem-se que $k(\mathbf{x}, \mathbf{x}) = 1$, conseqüentemente o cálculo da distância usando núcleos se reduz a: $d_k(\mathbf{x}, \mathbf{z}) = \sqrt{2 - 2k(\mathbf{x}, \mathbf{z})}$. Por esse fato, as funções núcleo podem ser interpretadas como uma medida de similaridade entre os pontos \mathbf{x} e \mathbf{z} .

Observe que quando é usado o núcleo euclidiano $k(\mathbf{x}, \mathbf{z}) = \langle \mathbf{x}, \mathbf{z} \rangle$, ou seja $\phi(\mathbf{x}) = \mathbf{x}$, a distância $d_k(\mathbf{x}, \mathbf{z})$ é a própria distância Euclidiana entre \mathbf{x} e \mathbf{z} no espaço \mathbb{R}^n , que será denotada por $\delta(\mathbf{x}, \mathbf{z})$.

Para exemplificar o uso das funções d_k para diferentes núcleos foram gerados dois gráficos, um para o caso unidimensional e outro para o caso bidimensional. Essas figuras representam uma visualização da função $d_k(\mathbf{x}, \mathbf{0})$, fazendo variar os valores de \mathbf{x} no intervalo $[-1, 1]$ para o primeiro caso e no quadrado $[-1, 1] \times [-1, 1]$ para o segundo caso. No caso bidimensional foi escolhida uma visualização por preenchimento de regiões entre as curvas de nível de $d_k(\mathbf{x}, \mathbf{0})$.

Para esses exemplos foram usados os seguintes núcleos: euclidiano (Figuras 2.1 e 2.2), gaussiano (Figuras 2.3 e 2.4), wave (Figuras 2.5 e 2.6), wavelets (Figuras 2.7 e 2.8), circular (Figuras 2.9 e 2.10), exponencial (Figuras 2.11 e 2.12), polinomial (Figuras 2.13 e 2.14) e o racional quadrático (Figuras 2.15 e 2.16).

É possível notar visualmente que existem diferenças entre todas essas funções distâncias. E é por esse fato que esse trabalho propõe uma simples mudança no método de Shepard, que usa o cálculo de distâncias como elemento básico do método de interpolação.

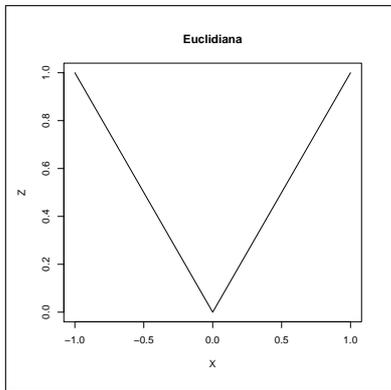


Figura 2.1: Gráfico da função distância usando o núcleo euclidiano unidimensional.

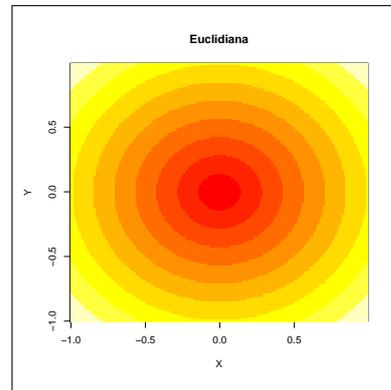


Figura 2.2: Gráfico da função distância usando o núcleo euclidiano bidimensional.

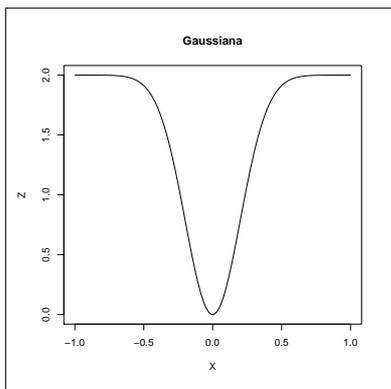


Figura 2.3: Gráfico da função distância usando o núcleo gaussiano unidimensional com parâmetro $\theta = 0.20$.

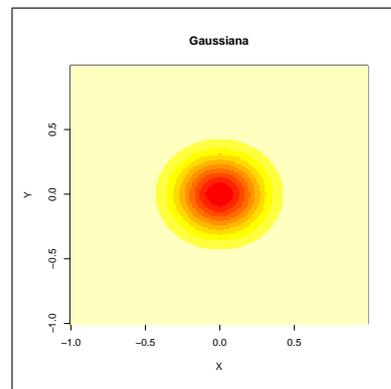


Figura 2.4: Gráfico da função distância usando o núcleo gaussiano bidimensional com parâmetro $\theta = 0.20$.

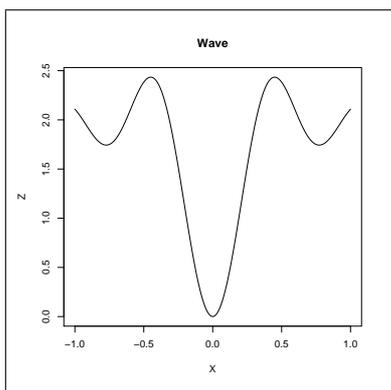


Figura 2.5: Gráfico da função distância usando o núcleo wave unidimensional com parâmetro $\theta = 0.10\pi$.

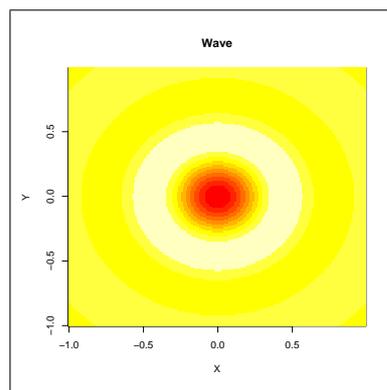


Figura 2.6: Gráfico da função distância usando o núcleo wave bidimensional com parâmetro $\theta = 0.10\pi$.

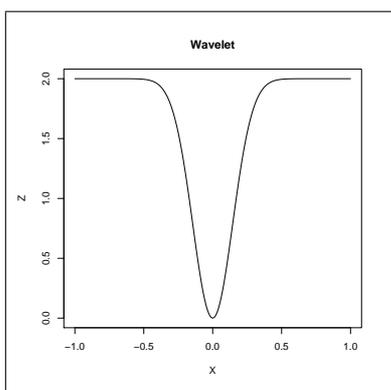


Figura 2.7: Gráfico da função distância usando o núcleo wavelets unidimensional com parâmetro $\theta = 0.15$.

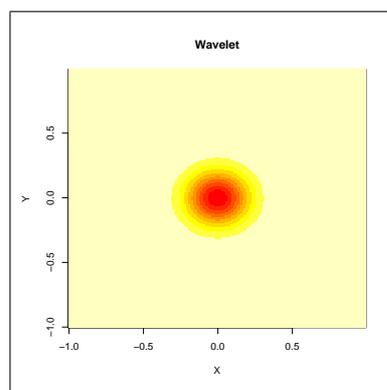


Figura 2.8: Gráfico da função distância usando o núcleo wavelets bidimensional com parâmetro $\theta = 0.15$.

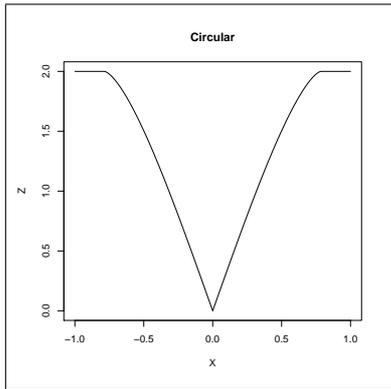


Figura 2.9: Gráfico da função distância usando o núcleo circular unidimensional com parâmetro $\theta = \frac{\pi}{4}$.

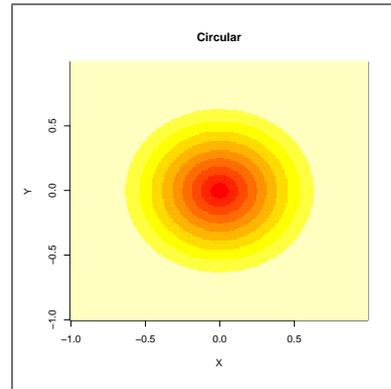


Figura 2.10: Gráfico da função distância usando o núcleo circular bidimensional com parâmetro $\theta = \frac{\pi}{4}$.

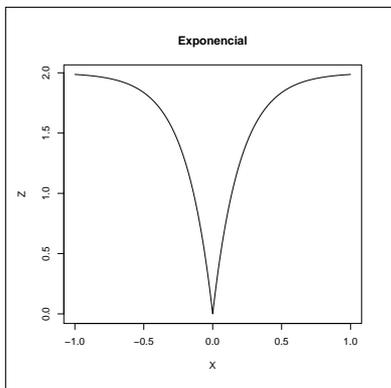


Figura 2.11: Gráfico da função distância usando o núcleo exponencial unidimensional com parâmetro $\theta = 0.2$.

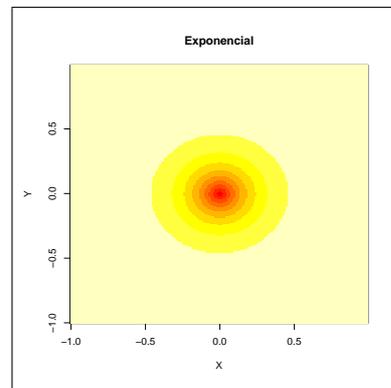


Figura 2.12: Gráfico da função distância usando o núcleo exponencial bidimensional com parâmetro $\theta = 0.2$.

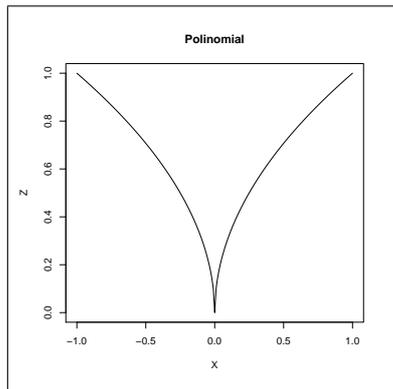


Figura 2.13: Gráfico da função distância usando o núcleo polinomial unidimensional com parâmetro $\theta = 0.25$.

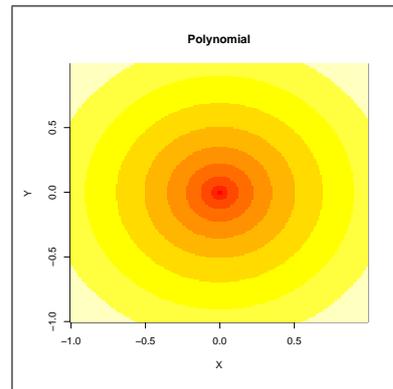


Figura 2.14: Gráfico da função distância usando o núcleo polinomial bidimensional com parâmetro $\theta = 0.25$.

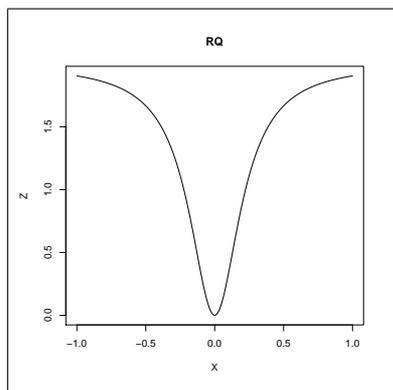


Figura 2.15: Gráfico da função distância usando o núcleo RQ unidimensional com parâmetro $\theta = 0.05$.

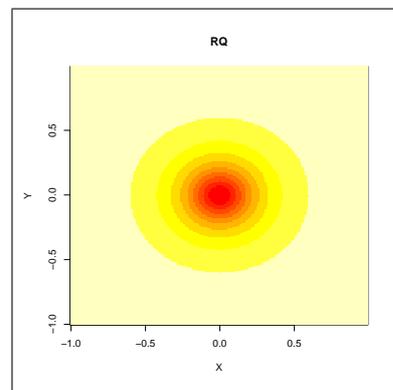


Figura 2.16: Gráfico da função distância usando o núcleo RQ bidimensional com parâmetro $\theta = 0.05$.