KDE

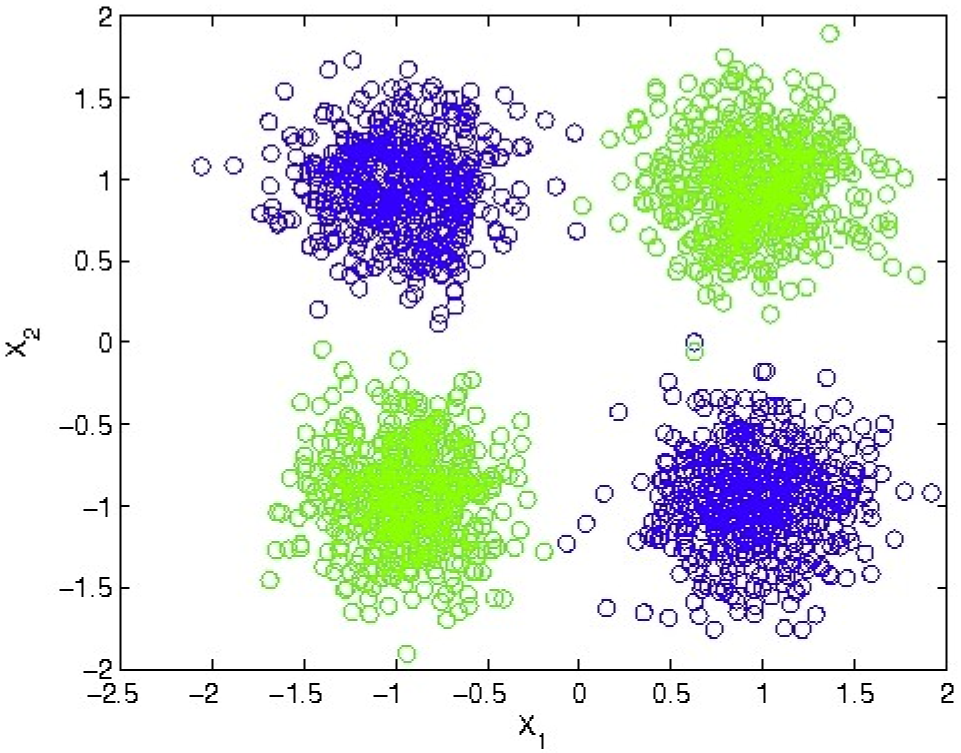
Estimador por densidade de kernel

Prof. Frederico Coelho.

(Baseado no livro de Antônio de Pádua Braga)

Introdução

Amostras atendem critério de normalidade



Distribuição pode ser estimada por:
- Média;
- Desvio-padrão;
- Correlação.

KDE

Amostras com distribuição não tão bem comportadas

- Modelos normais não se aplicam

Possível estimar função de densidade de probabilidade a partir de modelos não-paramétricos como o KDE

KDE

A estimativa pelo KDE se dá através da superposição de funções de densidade em cada ponto da amostra.

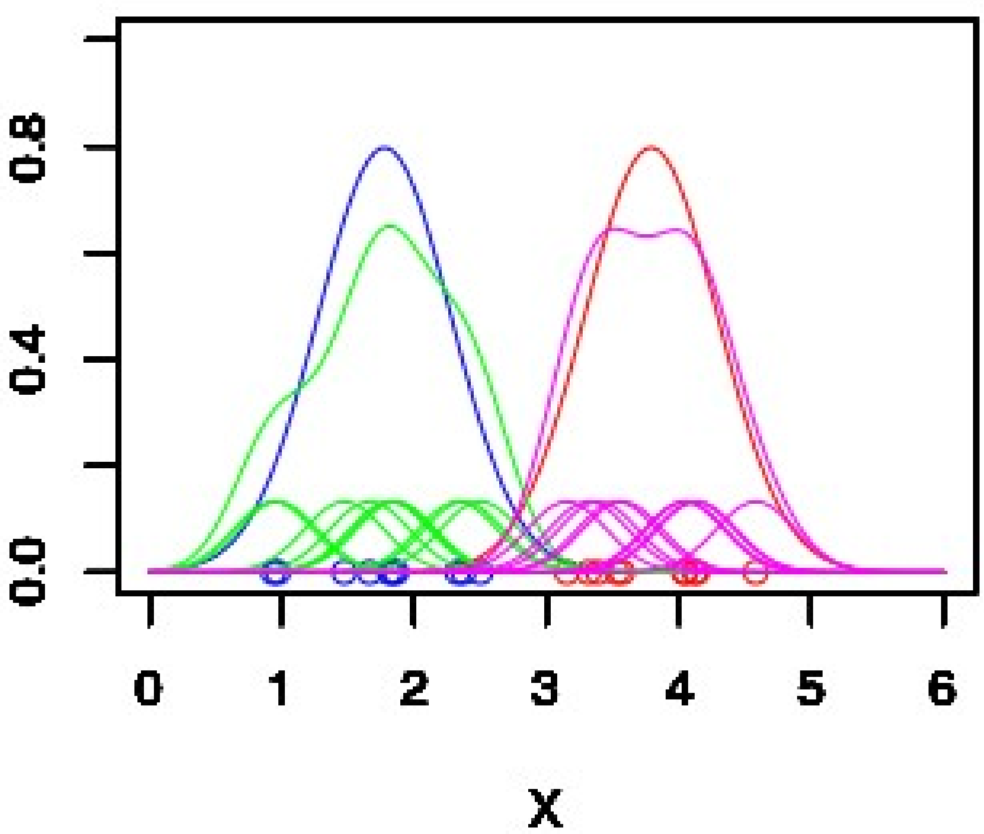
A restrição é que esta função seja simétrica e com integral unitária.

Típicamente usa-se a função de densidade normal como função de kernel

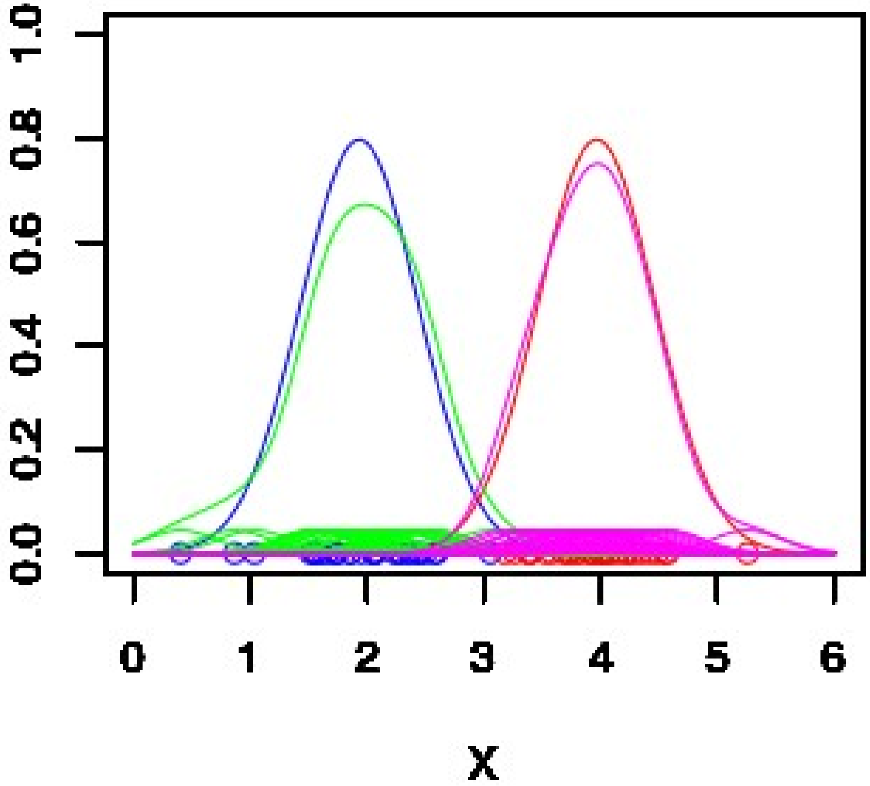
$$p(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h}K\left(\frac{x-x_i}{h}\right)$$

$$K\left(\frac{x-x_i}{h}\right) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-x_i}{h}\right)^2}$$

KDE



KDE



KDE

Para modelagens multidimensionais esta é uma abordagem interessante uma vez
Que o projetista não precisa definir o número de gaussianas, apenas a abertura *h* da função

Existem diversos métodos para definir *h* na literatura
- Sua definição é crítica e afeta muito o resultado final
- Deve ser estimado diretamente das amostras

Regra prática de Silverman:

$$h \approx 1.06\hat{\sigma}N^{-\frac{1}{5}}$$

KDE multivariado

Construir estimador ingênuo: assume-se independência entre variáveis.

Densidade resultante é o produto das densidade de cada variável.

$$p(\mathbf{x}_i) = \frac{1}{N}\sum_{j=1}^{N} \frac{1}{h}K\left(\frac{x_{i1}-x_{j1}}{h}\right)\frac{1}{h}K\left(\frac{x_{i2}-x_{j2}}{h}\right)\cdots\frac{1}{h}K\left(\frac{x_{in}-x_{jn}}{h}\right)$$

$$p(\mathbf{x}_i) = \frac{1}{N}\sum_{j=1}^{N} \frac{1}{\sqrt{2\pi}h}e^{-\frac{1}{2}\left(\frac{x_{i1}-x_{j1}}{h}\right)^2}\frac{1}{\sqrt{2\pi}h}e^{-\frac{1}{2}\left(\frac{x_{i2}-x_{j2}}{h}\right)^2}\cdots\frac{1}{\sqrt{2\pi}h}e^{-\frac{1}{2}\left(\frac{x_{in}-x_{jn}}{h}\right)^2}$$

$$p(\mathbf{x}_i) = \frac{1}{N(\sqrt{2\pi}h)^n}\sum_{j=1}^{N} e^{-\frac{(\mathbf{x}_i-\mathbf{x}_j)^2}{2h^2}}$$