

# Kernel Density Estimation (KDE)

## Classification Algorithms Seminar

Johannes Gätjen

29.01.2016

# Inhalt

Warum Histogramme uncool sind

Klassifikation mit KDE

Bandbreitenbestimmung

Implementation

Vor- und Nachteile

# Bandbreitenbestimmung

Verschiedene Verfahren möglich

1. Manuell, Trial and Error
2. Analytische Lösung per Annahme über Verteilung  
z.B. Silverman's rule of thumb:

$$\left( \frac{4\sigma^5}{3n} \right)^{\frac{1}{5}}$$

3. Iterative Verfahren, um Fehlermaß zu minimieren

## Implementationsdetails

- ▶ Verschiedene Kernel und Metriken möglich
- ▶ Bandbreite manuell, Silverman oder plug-in
- ▶ Bandbreite für alle Dimensionen einheitlich

## Implementationsdetails

- ▶ Verschiedene Kernel und Metriken möglich
- ▶ Bandbreite manuell, Silverman oder plug-in
- ▶ Bandbreite für alle Dimensionen einheitlich

Performance:

Iris	Seeds	Spirals	Mushrooms	Digits
94.5%	94.2%	100%	0%	0%

## Vor- und Nachteile

### Vorteile:

- ▶ Wenige Annahmen über Daten
  - ▶ jede Verteilung möglich (konkave Klassen etc.)
  - ▶ unterschiedliche Klassengrößen
  - ▶ unterschiedliche Misclassification costs
- ▶ Training in  $O(1)$
- ▶ Bayes-optimal

### Nachteile:

- ▶ relativ hohe Klassifikationszeit,  $O(N)$
- ▶ Bandbreitenbestimmung meist aufwändig
- ▶ vorhandenes Vorwissen nicht nutzbar

Ende