

Parzen Window Klassifikator

Überblick und Einführung

Florian Bethe, Angelika Ophagen

16. Januar 2017



FAKULTÄT FÜR
INFORMATIK

Inhaltsverzeichnis

Motivation

Methoden

Zusammenfassung

Literatur

Motivation

- ▶ Annahme: Daten wurden von unterliegendem Modell erzeugt (generativ)
- ▶ Modell ist unbekannt
- ▶ Parametrische Klassifikatoren brauchen Parameter, welche ermittelt werden müssen
- ▶ Klassenwahrscheinlichkeiten werden gelegentlich benötigt (z.B. Active Learning)

Kerndichteschätzung

- ▶ Annahme: Daten wurden von Modell generiert
- ▶ Gegeben Trainingsdaten: wie wahrscheinlich ist ein Datenpunkt?
- ▶ $\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)$
- ▶ K ist die Kernelfunktion (für gewöhnlich Gauß)
- ▶ h ist die *Bandbreite*

Bandbreite

- ▶ Beeinflusst den Grad der Glättung des Kernels
- ▶ Optimale Bandbreite nicht geschlossen berechenbar

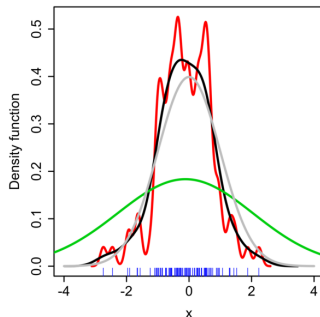


Abbildung 1: Vergleich verschiedener Bandbreiten

Silverman's Rule of Thumb

- ▶ Empirisch ermittelte Faustregel für Gaußkern
- ▶ Basiert auf geschätzter Varianz
- ▶ Nur Diagonalmatrix!

Silverman's Rule of Thumb

- ▶ Empirisch ermittelte Faustregel für Gaußkern
- ▶ Basiert auf geschätzter Varianz
- ▶ Nur Diagonalmatrix!

Zum Beispiel:

$$b(x) = \hat{\sigma}(x, x) \cdot \frac{4}{(d+2)n} \frac{1}{d+4}$$

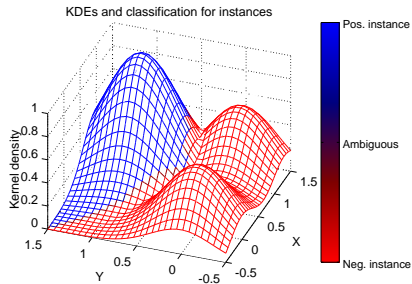
($n = \#$ Instanzen, $d = \#$ Dimensionen)

Monte Carlo Markov Chain

- ▶ Nutze Bayes: $p(h|x_1, \dots, x_n) \propto p(h) \cdot \mathcal{L}(x_1, \dots, x_n|h)$
- ▶ Likelihood mit Leave-one-out Schätzer
- ▶ Problem: Normalisierungskonstante nötig für direktes Sampling
- ▶ Ziehe Samples mittels Metropolis-Hastings
 - ▶ Generiere neue Samples mittels Random Walk (Markov-Kette)
 - ▶ Samples können abgelehnt werden, wenn sie zu unwahrscheinlich sind
 - ▶ Mittelung der Samples gibt Schätzung der optimalen Bandbreite

Parzen-Window-Klassifikator

- ▶ Nutze KDE zur Klassifikation
- ▶ Für jede Klasse separate KDE
- ▶ $p(c|x) \propto \hat{f}_h^c(x) \cdot \theta_c$
- ▶ Verschiedene Kernel pro Klasse möglich



Zusammenfassung & Ausblick

- ▶ 'Nichtparametrisch' ist ein wenig irreführend
- ▶ Bandbreite muss vorsichtig ermittelt werden
- ▶ Forschung an 'self-consistent' KDE, um dies zu vermeiden

Literatur I

- ▶ Parzen, E. (1962). On Estimation of a Probability Density Function and Mode
- ▶ Zhang, X., King, M.L., Hyndman, R.J. (2004). A Bayesian Approach to Bandwidth Selection for Multivariate Kernel Density Estimation
- ▶ Silverman, B.W. (1986). Density estimation for statistics and data analysis
- ▶ de Valpine, P. (2004). Monte Carlo State-Space Likelihoods by Weighted Posterior Kernel Density Estimation
- ▶ Buch-Kromann, T. (2007). Nonparametric kernel density estimation

Parzen Window Klassifikator

Überblick und Einführung

Florian Bethe, Angelika Ophagen

16. Januar 2017



FAKULTÄT FÜR
INFORMATIK