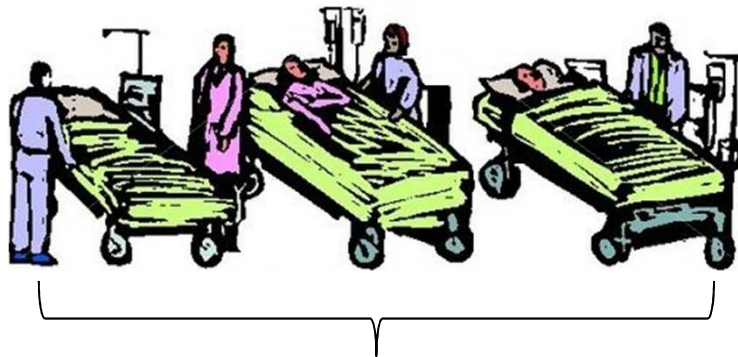


# Obiettivo: Classificare per Predire

- Predire lo stato renale finale del paziente alla fine della sua degenza in terapia intensiva (UCI)
- Punto di partenza: dati clinici e situazione del paziente al termine del primo giorno in UCI



Parametri alle 24 ore



Predizione della  
funzione renale  
all'uscita dall'UCI

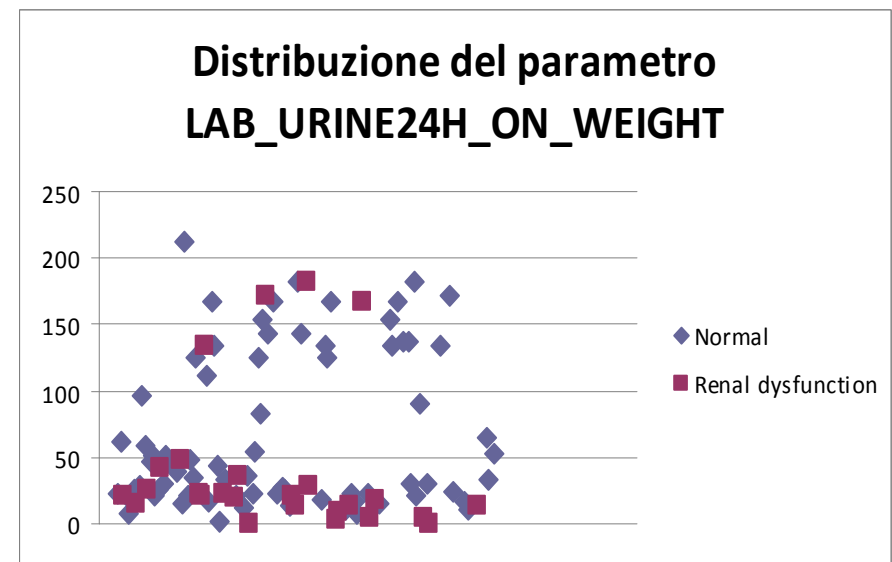
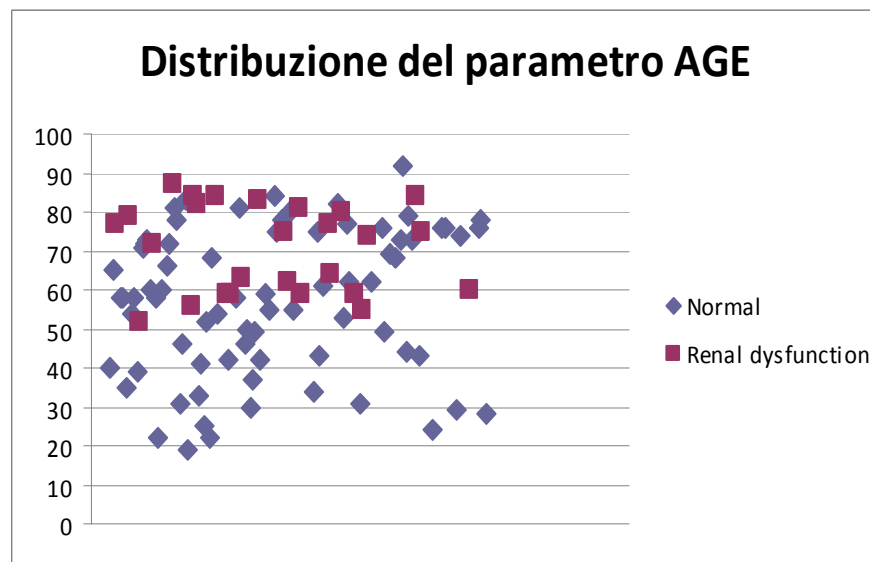
# Database: il progetto Nefroint

- Sviluppato nell'U.O. di Nefrologia, ULSS 6 (VI)
- Raccoglie dati da 23 centri di UCI italiani
- 100 pazienti con 47 features divise in:
  - Caso clinico e dettagli (18)
  - Punteggi dello score “APACHE II” (12)
  - Parametri da analisi di laboratorio (14)
  - Insufficienza Renale Acuta (IRA; AKI in inglese) (2)
  - Sepsì (1)

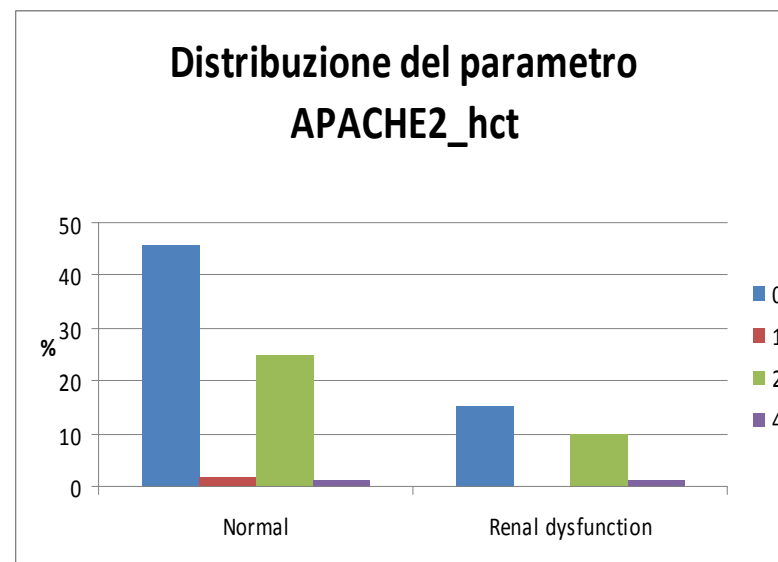
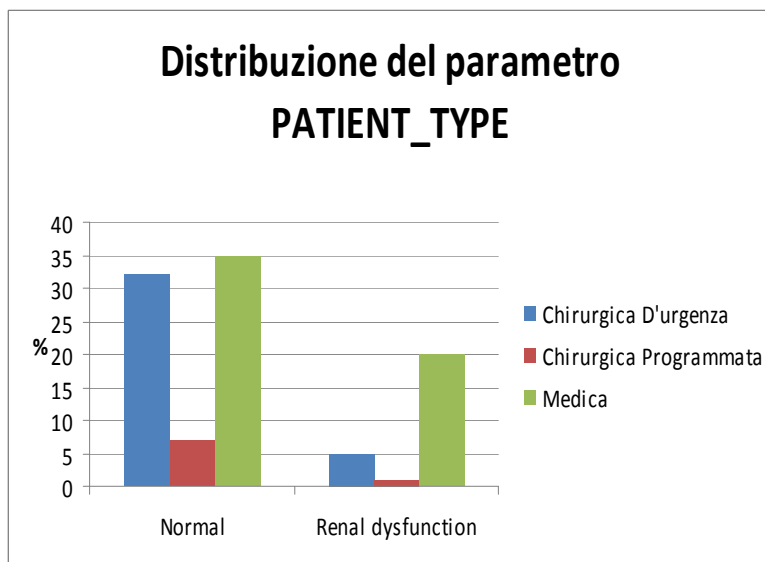
# Database: il training set

- Parte del progetto Nefroint
- 100 pazienti con 10 features:
  - Età
  - Tipologia del paziente all'ingresso
  - Diabete
  - Punteggio dello score "APACHE (Acute Physiology and Chronic Health Evaluation)" per Na, ematocrito, leucociti
  - Diuresi alle 24 ore e la peggiore delle 6 ore normalizzate al peso del paziente
  - Uso di diuretici
  - Insufficienza Renale Acuta nelle prime 24 ore
- 2 classi:
  - Normal (74): funzione renale normale
  - Renal dysfunction (26): funzionalità renale compromessa

## Alcuni parametri sono continui....



## Altri parametri sono discreti....



# Il software "Rapidminer"

- Precedentemente chiamato Y.A.L.E. (Yet Another Learning Environment)
- Implementa una vasta gamma di librerie:
  - Pre-processing dei dati
  - Classificatori supervisionati e non
  - Feature Selection
  - Validazione
  - ...
- Disponibile sul sito: [www.rapidminer.com](http://www.rapidminer.com) con una serie di tutorial su You Tube

Register with RapidMiner x RapidMiner Tutorial (p... x

https://www.youtube.com/watch?v=8FOZnQhKpCs

App DIPARTIMENTO DI I... Bacheche DEI e siti ... Facoltà di Ingegneri... News Università degli Stu... vocabolario - Dizion... www.ing.unipd.it/Pr... » Altri Preferiti

YouTube IT

Carica

Prossimo video Riproduzione automatica

**RapidMiner Tutorial (part 7/9) Naïve Bayes Classification**

RapidMinerTutorial

14.197

Aggiungi a Condividi Altro

Caricato il 29 giu 2011

This tutorial starts with introduction of Dataset. All aspects of dataset are discussed. Then basic working of RapidMiner is discussed. Once the viewer is acquainted with the knowledge of dataset and basic working of RapidMiner, following operations are performed on the dataset.

MOSTRA ALTRO

TUTTI I COMMENTI (9)

Condividi le tue opinioni

Commenti più popolari

**Prossimo video**

**RapidMiner Tutorial (part 8/9) Decision Tree (2 of 2)**  
di RapidMinerTutorial  
6.393 visualizzazioni

**RapidMiner Tutorial (part 6/9) K-NN Classification**  
di RapidMinerTutorial  
13.355 visualizzazioni

**RapidMiner Tutorial (part 9/9) Association Rules**  
di RapidMinerTutorial  
13.001 visualizzazioni

**Bayes Classifiers (1)**  
di Alexander Ihler  
7.174 visualizzazioni

**Bayes Classifiers (2): Naive Bayes**  
di Alexander Ihler  
6.841 visualizzazioni

**Twitter Data Sentiment Analysis Using RapidMiner**  
di Martin M  
17.535 visualizzazioni

**Data Visualisation - Part 1 using RapidMiner**  
di Markus Hofmann  
8.181 visualizzazioni

**Weka Tutorial 09: Feature Selection with Wrapper (Data Dimensionality)**  
di Rushdi Shams  
21.295 visualizzazioni

**Predicting the Stock Value using RapidMiner**  
di djshaily  
9.072 visualizzazioni

**5H K-Means Clustering in RapidMiner**  
di Dave Munro

11:54 27/05/2015

## Tra i metodi di classificazione:

- Classificatori Bayesiani
- Reti neurali
- Alberi decisionali
- Classificatori k-NN

## Tra i metodi di selezione:

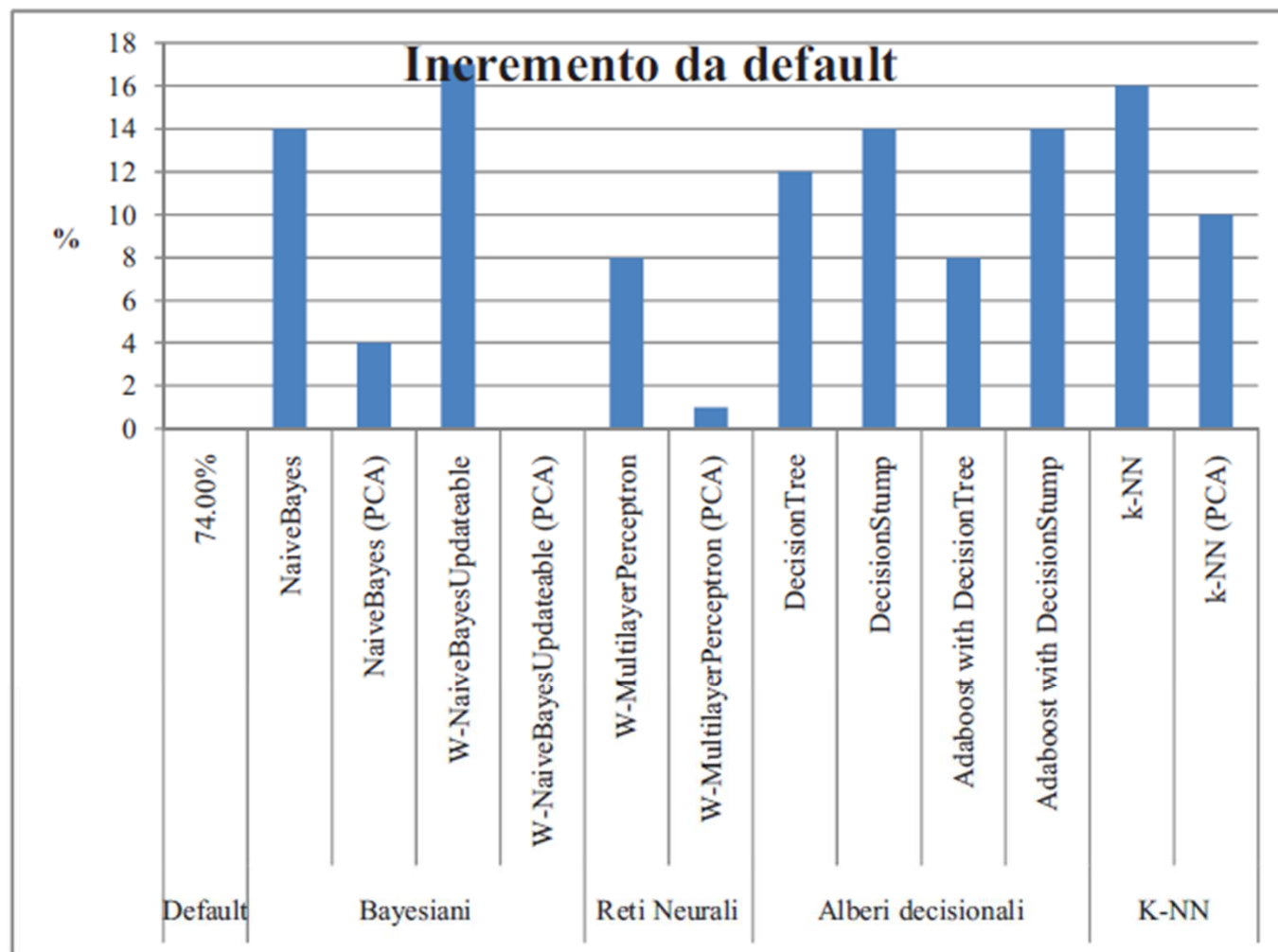
- Forward selection
- Backword elimination

## Tra i metodi di validazione:

- Leave-one-out cross validation



# RISULTATI



# Classificatore W-Naive Bayes

## Tabella di contingenza

		True			
		Renal dysfunction	Normal		
Predicted	Renal dysfunction	18	1	94.74%	ppv
	Normal	8	73	90.12%	npv
		69.23%	98.65%		
		sensitivity	specificity		

I parametri selezionati sono:

- APACHE2\_Creatinine\*
- lab\_hemoglobin
- lab\_WBC
- AKI\_first\_day
- sepsis\_type\_id
- rrt
- APACHE2\_Na
- Weight
- patient\_type
- APACHE2\_temp
- APACHE2\_MAP
- lab\_diuretic\_use

# Si possono definire vari indici

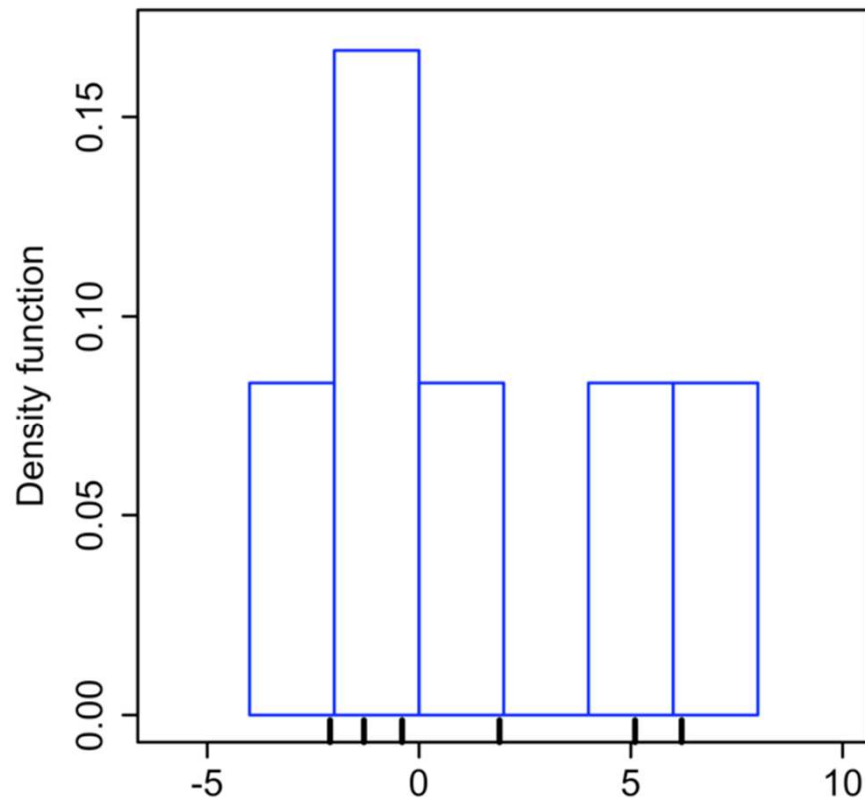
		Condition (as determined by "Gold standard")			
Total population		Condition positive	Condition negative	Prevalence = $\frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	
Test outcome	Test outcome positive	True positive	False positive (Type I error)	Positive predictive value (PPV, Precision) = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Test outcome positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Test outcome positive}}$
	Test outcome negative	False negative (Type II error)	True negative	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Test outcome negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Test outcome negative}}$
Positive likelihood ratio (LR+) = TPR/FPR		True positive rate (TPR, Sensitivity, Recall) = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR, Fall-out) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Negative likelihood ratio (LR-) = FNR/TNR		False negative rate (FNR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	True negative rate (TNR, Specificity, SPC) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$		
Diagnostic odds ratio (DOR) = LR+/LR-					

# Si può generalizzare al caso di più classi : Confusion Table

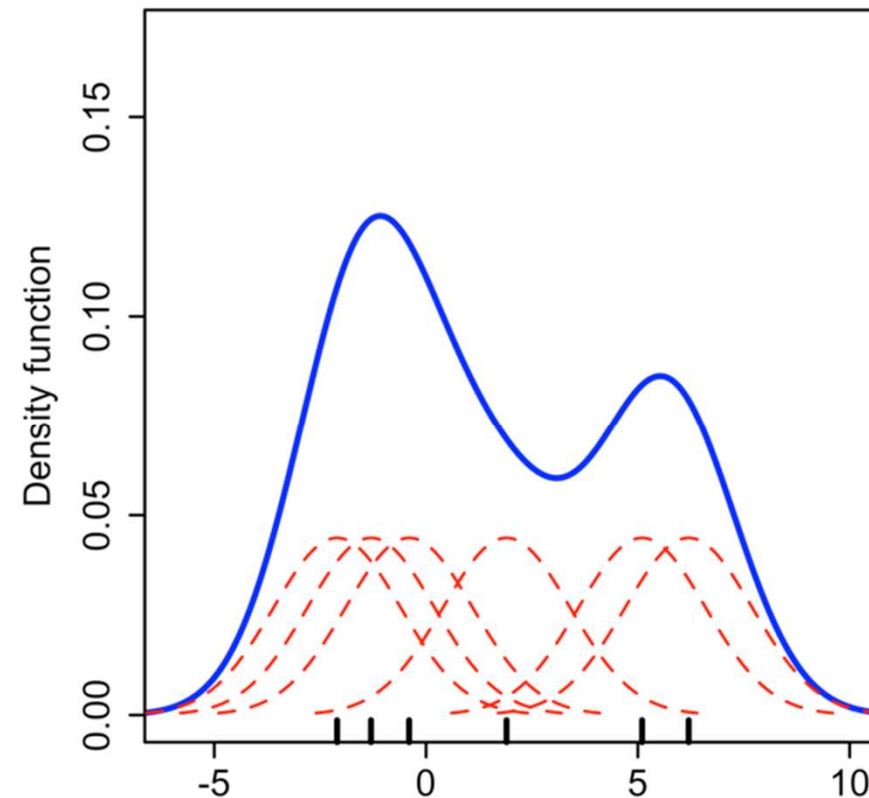
Accuracy: 84.59%		TRUE			
		Class 1	Class 2	Class 3	class precision
Predicted	Class 1	377	19	26	89.34%
	Class 2	15	62	31	57.41%
	Class 3	25	37	400	86.58%
	class recall	90.41%	52.54%	87.53%	

# Stima della densità di probabilità

istogramma

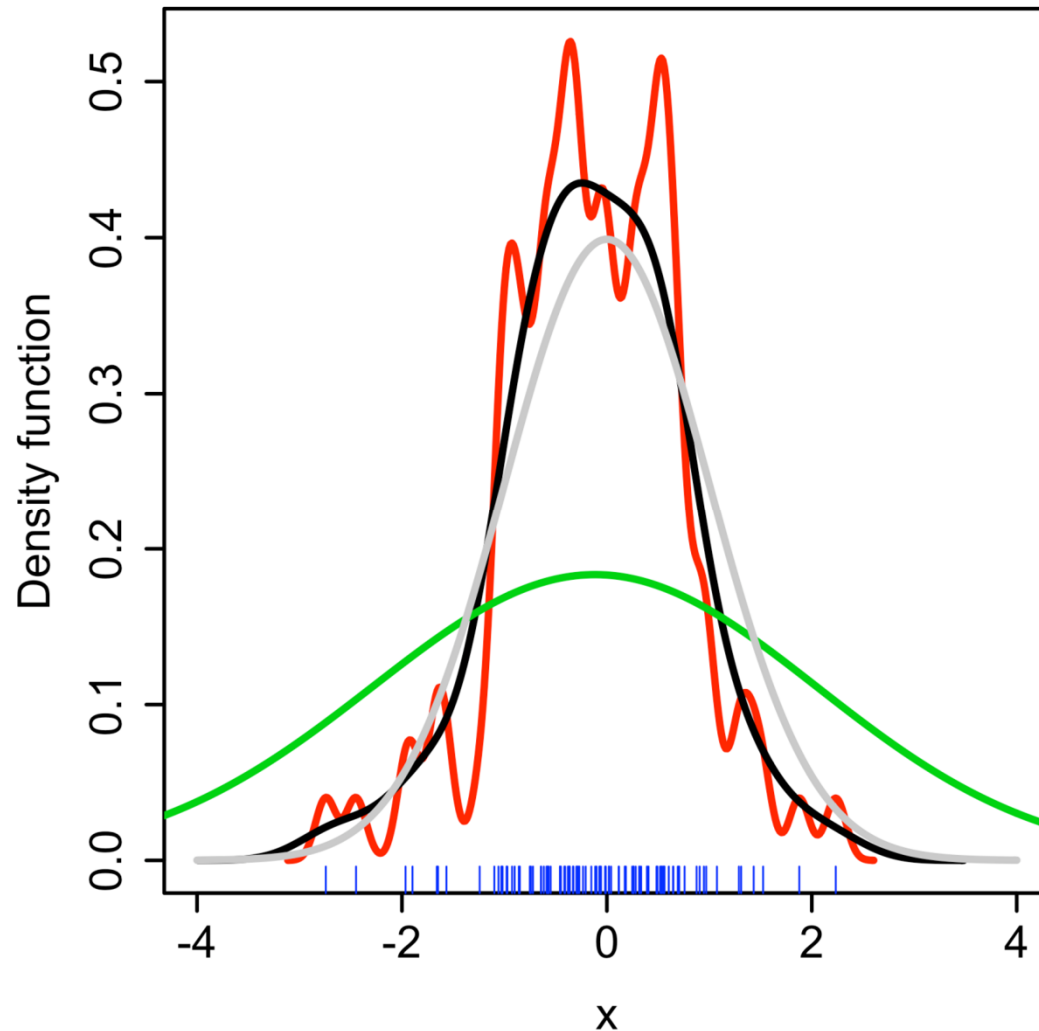


kernel



For the kernel density estimate, we place a normal kernel with variance 2.25 (red dashed lines) on each of the data points  $x_i$ . The kernels are summed to make the kernel density estimate (solid blue curve). The smoothness of the kernel density estimate is evident compared to the discreteness of the histogram, as kernel density estimates converge faster to the true underlying density for continuous random variables

# Stima con kernels



Kernel density estimate with different bandwidths of a random sample of 100 points from a standard normal distribution. Grey: true density (standard normal). Red: KDE with  $h=0.05$ . Black: KDE with  $h=0.337$ . Green: KDE with  $h=2$ .