LES FACULTÉS
DE L'UNIVERSITÉ
CATHOLIQUE DE LILLE

Non Parametric Models Examples

NON-PARAMETRIC STATISTICS

Baptiste Mokas

baptiste.mokas@gmail.com

weeki.io/dynamical-système

linktr.ee/baptistemokas

+33 7 69 08 54 19

KNOWLEDGE TREE

Part 1: Introduction to Nonparametric Models Examples

1.1 Kernel Density Estimation

- Kernel Functions and Bandwidth Selection
- Density Estimation for Continuous Data
- Applications in Probability Density Estimation

1.2 Nonparametric Regression

- Local Regression (LOESS)
- Nadaraya-Watson Kernel Regression
- Smoothing Techniques for Noisy Data

1.3 Nonparametric Classification

- k-Nearest Neighbors (k-NN)
- Decision Trees for Classification
- Support Vector Machines (SVM) with Nonlinear Kernels

Part 2: Nonparametric Statistics Fundamentals

2.1 Rank-Order Statistics

- Rank Tests (Wilcoxon Signed-Rank Test, Mann-Whitney U Test)
- Spearman's Rank Correlation
- Kendall's Tau Rank Correlation

2.2 Bootstrap Resampling

- Bootstrap Methodology
- Bootstrapping Confidence Intervals
- Bootstrap for Hypothesis Testing

2.3 Nonparametric Tests for Independence

- Chi-Square Test for Independence
- Fisher's Exact Test
- Nonparametric Association Tests (Kruskal-Wallis, Friedman)

Part 3: Nonparametric Data Visualization and Exploration

3.1 Histograms and Smoothing

- Histogram Construction
- Kernel Density Plots
- Rug Plots and Probability Density Estimation

3.2 Nonparametric Exploratory Data Analysis (EDA)

- Box Plots and Violin Plots
- Nonparametric Outlier Detection
- Nonparametric Hypothesis Tests for Location and Scale

3.3 Nonparametric Regression Plots

- Scatterplot Smoothers
- Locally Weighted Scatterplot Smoothing (LOWESS)
- Nonparametric Regression in Data Exploration

Part 4: Applications of Nonparametric Statistics

4.1 Biomedical Data Analysis

- Nonparametric Survival Analysis (Kaplan-Meier)
- Rank-Based Tests in Clinical Research
- Nonparametric Regression in Epidemiology

4.2 Environmental Data Analysis

- Nonparametric Spatial Interpolation
- Nonparametric Analysis of Spatial Clusters
- Ecological Niche Modeling

4.3 Business and Finance

- Nonparametric Tests for Portfolio Performance
- Nonparametric Forecasting Models
- Nonparametric Methods in Risk Assessment

KEYWORDS (NEW)

Fenêtre de lissage
Estimateur de Nadaraya-Watson
ANOVA de Friedman
Noyau quadratique
Test non paramétrique
Test du signe
Noyau d'Epanechninov
Régression non paramétrique
Modèle additif
Noyau rectangulaire
Test de Kruskal-Wallis
Régressogramme
Régressogramme mobile
Tests statistiques non paramétriques
Noyau gaussien
Noyau uniforme

ADD TO TREE

11 Tests non paramétriques
11.1 Test de Mann-Whitney
11.2 Test de Wilcoxon (Wilcoron signed rank test)
11.3 Test de Corrélation de rang de Spearman A Table de Mann-Whitney
B Table de Wilcoxon
C Table du cofficient de rang de Spearman

Chapitre 6
Fiche 78 Les tests non paramétriques pour analyses univariées
Fiche 79 Le test U de Mann-Whitney
Fiche 80 Le test de Kruskal-Wallis
Fiche 81| Le test T de Wilcoxon
Fiche 82 Le test de Friedman Quels tests post hoc utiliser après un test sur les rangs
Fiche 83
Fiche 84 Le test du z sur table de contingence
Fiche 85 Le calcul de probabilité exacte (CPE) de Fisher Comment comparer une proportion observée
Fiche 86 à une proportion théorique
Fiche 87 Comment comparer plusieurs proportions indépendantes
Comment comparer deux proportions en échantillons appariés : le test de McNemar
Fiche 88 Comment comparer plus de deux proportions en echantilions apparies : le test Q de Cochran
Fiche 89 Comment comparer deux distributions empiriques: le test de Kolmogorov-Smirnov
Fiche 90 Comment comparer une distribution empirique • une distribution théorique
Fiche 91 Les tests d'asymétrie et d'aplatissement Focus QCM Les tests statistiques à l'épreuve des tests statistiques : crash test

KEYWORDS

- Nonparametric Modeling
- Kernel Density Estimation
- Probability Density Estimation
- Nonparametric Regression
- Local Regression (LOESS)
- Nadaraya-Watson Kernel Regression
- Nonparametric Classification
- k-Nearest Neighbors (k-NN)
- Decision Trees
- Support Vector Machines (SVM)
- Rank-Order Statistics
- Wilcoxon Signed-Rank Test
- Mann-Whitney U Test
- Bootstrap Resampling
- Data Visualization
- Exploratory Data Analysis (EDA)
- Histograms
- Kernel Density Plots
- Rug Plots
- Biomedical Data Analysis
- Environmental Data Analysis
- Business and Finance

Use Case with python

Use Case: Nonparametric Analysis of Environmental Data

In the context of the course on "Nonparametric Models Examples," "Nonparametric Statistics Fundamentals," "Nonparametric Data Visualization and Exploration," and "Applications of Nonparametric Statistics," which covers topics related to kernel density estimation, nonparametric regression, nonparametric tests, data visualization, and various applications, let's explore a use case related to the nonparametric analysis of environmental data.

Description:

In this use case, we will apply nonparametric statistical techniques to analyze environmental data, specifically focusing on air quality measurements in urban areas. We want to understand the spatial distribution of air pollutants and identify regions with high pollution levels.

Key Components:

Kernel Density Estimation: Understanding kernel density estimation and how it can be used to estimate the probability density function of air pollutant concentrations.

Nonparametric Data Visualization and Exploration: Creating kernel density plots and other nonparametric visualization techniques to visualize the spatial distribution of air pollutant concentrations.

Nonparametric Tests for Independence: Using nonparametric tests like the Kruskal-Wallis test to assess whether there are significant differences in pollutant levels across different geographical regions.

Nonparametric Spatial Interpolation: Applying nonparametric spatial interpolation techniques to estimate pollutant concentrations at unmonitored locations within the study area.

Environmental Data Analysis Scenario:

Imagine you work for a government agency responsible for monitoring air quality in a large urban area. You have access to a dataset containing air pollutant measurements (e.g., PM2.5, NO2, CO) collected from various monitoring stations across the city over time. Here's how you can apply nonparametric statistics to analyze this data:

Python Code Example (Nonparametric Analysis of Air Quality Data):

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy import stats
from sklearn.neighbors import KernelDensity
from sklearn.model_selection import GridSearchCV

# Load air quality data with latitude and longitude information
data = pd.read_csv('air_quality_data.csv')

# Visualize the spatial distribution of pollutant concentrations using kernel density estimation
X = data[['Latitude', 'Longitude']]
pollutant_data = data['PM2.5']  # Replace with the specific pollutant of interest
kde = KernelDensity(bandwidth=0.1, kernel='gaussian')
kde.fit(X)
grid_x, grid_y = np.meshgrid(np.linspace(X['Latitude'].min(), X['Latitude'].max(), 100),
                             np.linspace(X['Longitude'].min(), X['Longitude'].max(), 100))
grid_coords = np.vstack([grid_x.ravel(), grid_y.ravel()]).T
log_density = kde.score_samples(grid_coords)
density = np.exp(log_density).reshape(grid_x.shape)
plt.contourf(grid_x, grid_y, density, cmap='viridis')
plt.colorbar(label='Density')
plt.scatter(X['Latitude'], X['Longitude'], c=pollutant_data, cmap='Reds', s=10)
plt.xlabel('Latitude')
plt.ylabel('Longitude')
plt.title('Spatial Distribution of PM2.5 Concentrations')
plt.show()

# Perform a Kruskal-Wallis test to assess differences in pollutant levels across regions
regions = data['Region']  # Replace with the specific region variable
statistic, p_value = stats.kruskal(*[pollutant_data[regions == region] for region in regions.unique()])
print(f'Kruskal-Wallis Statistic: {statistic}')
print(f'P-value: {p_value}')

# Apply nonparametric spatial interpolation to estimate pollutant concentrations at unmonitored locations
# (e.g., using inverse distance weighting)

```

In this code, we load air quality data, visualize the spatial distribution of pollutant concentrations using kernel density estimation, perform a Kruskal-Wallis test to assess regional differences, and apply nonparametric spatial interpolation techniques to estimate pollutant concentrations at unmonitored locations.

This use case demonstrates how nonparametric statistics can be employed to analyze environmental data, identify pollution hotspots, and make informed decisions for urban planning and public health management.

REFERENCES V2

TOP 20 REFERENCES

- Hollander, M., Wolfe, D. A., & Chicken, E. (2014). Nonparametric Statistical Methods. Wiley.
- Wasserman, L. (2006). All of Nonparametric Statistics. Springer.
- Efron, B., & Tibshirani, R. J. (1994). An Introduction to the Bootstrap. Chapman & Hall.
- Hastie, T., & Tibshirani, R. (1990). Generalized Additive Models. Chapman & Hall.
- Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. CRC Press.
- Bowman, A. W., & Azzalini, A. (1997). Applied Smoothing Techniques for Data Analysis. Oxford University Press.
- Altman, D. G. (1991). Practical Statistics for Medical Research. Chapman and Hall.
- Manly, B. F. J. (2007). Randomization, Bootstrap and Monte Carlo Methods in Biology. Chapman & Hall.
- Simonoff, J. S. (1996). Smoothing Methods in Statistics. Springer.
- Conover, W. J. (1999). Practical Nonparametric Statistics. Wiley.
- Diggle, P. J., & Ribeiro Jr, P. J. (2007). Model-Based Geostatistics. Springer. (Nonparametric Spatial Interpolation)
- Lachenbruch, P. A. (2001). Comparisons of Two-Part Models with Competing Risks Models. Statistics in Medicine.
- Davison, A. C., & Hinkley, D. V. (1997). Bootstrap Methods and Their Applications. Cambridge University Press.
- Puri, M. L., & Sen, P. K. (1971). Nonparametric Methods in Multivariate Analysis. Wiley.
- Fox, J. (2000). Nonparametric Simple Regression: Smoothing Scatterplots. Sage Publications.
- Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern Classification. Wiley. (Nonparametric Classification Techniques)
- Gibbons, J. D., & Chakraborti, S. (2011). Nonparametric Statistical Inference. CRC Press.
- Lehmann, E. L., & D'Abrera, H. J. M. (2006). Nonparametrics: Statistical Methods Based on Ranks. Springer.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). The Elements of Statistical Learning. Springer. (Chapter on Nonparametric Methods)

TEXTE DE DESCRIPTION DU COURS

Dive into the realm of nonparametric modeling through this comprehensive course that delves into essential techniques and their real-world applications.

In this course, you'll unravel the intricacies of nonparametric modeling, a flexible approach that doesn't rely on strict distribution assumptions. The journey begins with a focus on fundamental techniques like kernel density estimation, which allows you to estimate probability density for continuous data, making it an essential tool in probability density estimation. You'll also explore nonparametric regression, including local regression (LOESS) and Nadaraya-Watson kernel regression, enabling you to model complex relationships in data without the constraints of parametric assumptions. Additionally, you'll delve into nonparametric classification techniques, including k-Nearest Neighbors (k-NN), decision trees, and support vector machines (SVM) with nonlinear kernels, all of which provide powerful tools for pattern recognition and classification tasks.

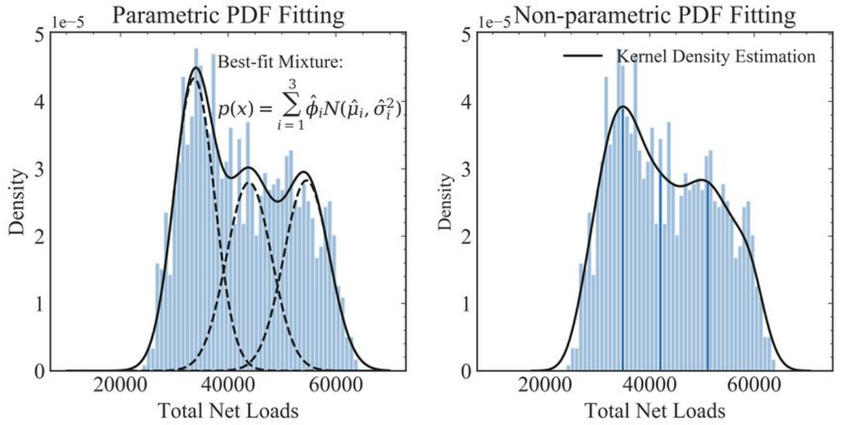
The course goes beyond modeling to uncover the fundamentals of nonparametric statistics. You'll delve into rank-order statistics, gaining proficiency in tests like the Wilcoxon Signed-Rank Test and Mann-Whitney U Test. These tools are essential for assessing relationships and differences when parametric assumptions don't hold. The course also introduces you to bootstrap resampling, a versatile technique that empowers you to estimate sampling distributions and construct confidence intervals without relying on specific distribution assumptions. You'll also learn how to leverage bootstrapping for hypothesis testing, enhancing your statistical toolkit.

Effective data visualization and exploration are integral aspects of nonparametric modeling. You'll discover how to construct informative histograms, create kernel density plots, and interpret rug plots to estimate probability densities accurately. The course also covers nonparametric exploratory data analysis (EDA), teaching you to use box plots, violin plots, and nonparametric hypothesis tests for location and scale to uncover patterns and outliers in your data.

As you progress through the course, you'll explore practical applications in biomedical data analysis, environmental data analysis, and business and finance. These real-world examples demonstrate how nonparametric modeling techniques can be applied to solve complex problems across diverse domains.

By the end of this course, you'll have a profound understanding of nonparametric modeling and its versatile applications, equipping you with valuable skills for data analysis and decision-making in various fields.

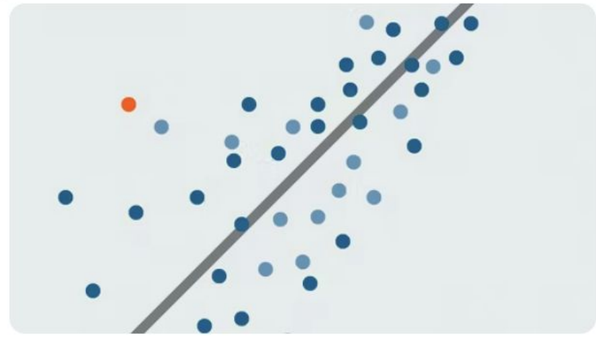
DESIGN DE LA CARD



Author: Baptiste Mokas, Weeki

Course Name: Simple Linear Regression

#NonparametricModeling
#StatisticalAnalysis
#DataExploration



Duke University

Linear Regression and Modeling

Compétences que vous acquerrez: Probability & Statistics, Regression, Business Analysis, Data Analysis, General Statistics, Statistical Analysis,…

4.8 (1.7k avis)

Débutant · Course · 1 à 4 semaines