ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ • ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING • UNIVERSITY OF IOANNINA

ΠΑΡΟΥΣΙΑΣΗ

ΔΙΔΑΚΤΟΡΙΚΗΣ ΔΙΑΤΡΙΒΗΣ

ΗΜΕΡΟΜΗΝΙΑ: Πέμπτη, 6 Φεβρουαρίου 2025

ΩΡΑ: 10.00 – 12:00

ΑΙΘΟΥΣΑ: Αίθουσα Σεμιναρίων ΤΜΗΥΠ

ΟΜΙΛΗΤΡΙΑ: Παρασκευή Χασάνη

Θ έ μ α

«Machine Learning Methods based on Unimodality Testing»

Επταμελής Εξεταστική Επιτροπή:

- **Αριστείδης Λύκας,** Καθηγητής του Τμήματος Μηχανικών Η/Υ & Πληροφορικής του Πανεπιστημίου Ιωαννίνων
- **Κωνσταντίνος Μπλέκας**, Καθηγητής του Τμήματος Μηχανικών Η/Υ & Πληροφορικής του Πανεπιστημίου Ιωαννίνων
- **Χριστόφορος Νίκου**, Καθηγητής του Τμήματος Μηχανικών Η/Υ & Πληροφορικής του Πανεπιστημίου Ιωαννίνων
- **Γεώργιος Στάμου**, Καθηγητής της Σχολής Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου
- **Γρηγόριος Τσουμάκας**, Καθηγητής του Τμήματος Πληροφορικής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης
- **Κωνσταντίνος Βλάχος**, Επίκουρος Καθηγητής του Τμήματος Μηχανικών Η/Υ & Πληροφορικής του Πανεπιστημίου Ιωαννίνων
- **Ιωάννης Παυλόπουλος**, Επίκουρος Καθηγητής του Τμήματος Πληροφορικής του Οικονομικού Πανεπιστημίου Αθηνών

Π ε ρ ί λ η ψ η:

Recognizing unimodal data distributions is of great significance in statistics, machine learning and data science. The characteristic property of a unimodal distribution is that data values are gathered around a single value (peak), which is

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

Τ.Θ. 1186, ΙΩΑΝΝΙΝΑ 45110
Τ: 265100 8817 - 8813 - 7196
http:/www.cse.uoi.gr/

DEPT. OF COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF IOANNINA

P.O. BOX 1186, IOANNINA
GR - 45110 GREECE
T: +30 265100 8817 - 8813 - 7196
http:/www.cse.uoi.gr/

the mode of the distribution. Due to this property, data can be characterized as homogeneous, forming a single and coherent group. Unimodality tests have been proposed to decide on the unimodality of a set of data values, thus providing useful knowledge about the structure of the data. This thesis concerns the development and implementation of machine learning methods based on the notion of unimodality.

At first a new unimodality test is proposed called Unimodal Uniform test (UU-test) to decide if a univariate dataset has been generated by a unimodal distribution or not. The method utilizes the empirical distribution function (ecdf) and attempts to obtain a unimodal piecewise linear approximation of the ecdf under the constraint that the data corresponding to each linear segment follow the uniform distribution. Compared to other unimodality tests, it also produces a generative model of the unimodal data in the form of a mixture of uniform distributions (UMM). Thus, it can be used for statistical modeling of data generated by unimodal distributions with arbitrary shape. Moreover, UMM performance is improved by substituting the uniform distribution with a more flexible and differential one, called Π-sigmoid. A mixture model of Π-sigmoids, called UΠsMM, is trained using a mechanism that maintains unimodality.

The problem of modeling univariate multimodal data is also addressed in this thesis, with two main contributions. At first, properties of critical points of the data ecdf are introduced that provide indications on the existence of density valleys. Using these properties, the UniSplit algorithm is proposed that detects valley points and partitions the dataset into unimodal subsets, automatically estimating their number. Next, a statistical model is proposed, called Unimodal Mixture Model (UDMM), which models each unimodal subset with a UMM. A key strength of UDMM is its flexibility and independence from specific parametric assumptions, making it well-suited for datasets generated by sources of different probability density (e.g., one Gaussian and one uniform). Another important property is that the number of components is automatically estimated, therefore, a major issue in mixture modeling is addressed.

Finally, an unsupervised method is proposed for clustering multidimensional data using decision trees. The concept of axis unimodal cluster is introduced, which is a cluster where all features are unimodal as decided by a unimodality test. A method is presented that constructs binary decision trees, providing axis-aligned partitions of the data and offering interpretable clustering solutions. Two criteria are proposed to identify the best split pair (feature and threshold) at each node, aiming to improve the unimodality of the partition after each split. Compared to other unsupervised decision tree methods, this approach has several advantages: it is simple, avoids preprocessing steps and does not employ computationally expensive optimization methods or difficult to tune hyperparameters, such as number of clusters or maximum tree depth.

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

Τ.Θ. 1186, ΙΩΑΝΝΙΝΑ 45110
Τ: 265100 8817 - 8813 - 7196
http:/www.cse.uoi.gr/

DEPT. OF COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF IOANNINA

P.O. BOX 1186, IOANNINA
GR - 45110 GREECE
T: +30 265100 8817 - 8813 - 7196
http:/www.cse.uoi.gr/