

1. Record Nr.	UNINA9910522959703321
Autore	Lespinats Sylvain
Titolo	Nonlinear Dimensionality Reduction Techniques : A Data Structure Preservation Approach / / by Sylvain Lespinats, Benoit Colange, Denys Dutykh
Pubbl/distr/stampa	Cham : , : Springer International Publishing : , : Imprint : Springer, , 2022
ISBN	3-030-81026-7
Edizione	[1st ed. 2022.]
Descrizione fisica	1 online resource (279 pages)
Disciplina	519.5
Soggetti	Machine learning Artificial intelligence - Data processing Image processing Machine Learning Data Science Image Processing
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	1 Data science context -- 1.1 Data in a metric space -- 1.1.1 Measuring dissimilarities and similarities -- 1.1.2 Neighbourhood ranks -- 1.1.3 Embedding space notations -- 1.1.4 Multidimensional data -- 1.1.5 Sequence data -- 1.1.6 Network data -- 1.1.7 A few multidimensional datasets -- 1.2 Automated tasks -- 1.2.1 Underlying distribution -- 1.2.2 Category identification -- 1.2.3 Data manifold analysis -- 1.2.4 Model learning -- 1.2.5 Regression -- 1.3 Visual exploration -- 1.3.1 Human in the loop using graphic variables -- 1.3.2 Spatialization and Gestalt principles -- 1.3.3 Scatter plots -- 1.3.4 Parallel coordinates -- 1.3.5 Colour coding -- 1.3.6 Multiple coordinated views and visual interaction -- 1.3.7 Graph drawing -- 2 Intrinsic dimensionality -- 2.1 Curse of dimensionality -- 2.1.1 Data sparsity -- 2.1.2 Norm concentration -- 2.2 ID estimation -- 2.2.1 Covariance-based approaches -- 2.2.2 Fractal approaches -- 2.2.3 Towards local estimation -- 2.3 TIDLE -- 2.3.1 Gaussian mixture modelling -- 2.3.2 Test of TIDLE on a two clusters case -- 3 Map

evaluation -- 3.1 Objective and practical indicators -- 3.1.1
Subjectivity of indicators -- 3.1.2 User studies on specific tasks -- 3.2
Unsupervised global evaluation -- 3.2.1 Types of distortions -- 3.2.2
Link between distortions and mapping continuity -- 3.2.3 Reasons of
distortions ubiquity -- 3.2.4 Scalar indicators -- 3.2.5 Aggregation --
3.2.6 Diagrams -- 3.3 Class-aware indicators -- 3.3.1 Class separation
and aggregation -- 3.3.2 Comparing scores between the two spaces --
3.3.3 Class cohesion and distinction -- 3.3.4 The case of one cluster
per class -- 4 Map interpretation -- 4.1 Axes recovery -- 4.1.1 Linear
case: biplots -- 4.1.2 Non-linear case -- 4.2 Local evaluation -- 4.2.1
Point-wise aggregation -- 4.2.2 One to many relations with focus point
-- 4.2.3 Many to many relations -- 4.3 MING -- 4.3.1 Uniform
formulation of rank-based indicator -- 4.3.2 MING graphs -- 4.3.3
MING analysis for a toy dataset -- 4.3.4 Impact of MING parameters --
4.3.5 Visual clutter -- 4.3.6 Oil flow -- 4.3.7 COIL-20 dataset -- 4.3.8
MING perspectives -- 5 Unsupervised DR -- 5.1 Spectral projections --
5.1.1 Principal Component Analysis -- 5.1.2 Classical MultiDimensional
Scaling -- 5.1.3 Kernel methods: Isomap, KPCA, LE -- 5.2 Non-linear
MDS -- 5.2.1 Metric MultiDimensional Scaling -- 5.2.2 Non-metric
MultiDimensional Scaling -- 5.3 Neighbourhood Embedding -- 5.3.1
General principle: SNE -- 5.3.2 Scale setting -- 5.3.3 Divergence
choice: NeRV and JSE -- 5.3.4 Symmetrization -- 5.3.5 Solving the
crowding problem: tSNE -- 5.3.6 Kernel choice -- 5.3.7 Adaptive
Student Kernel Imbedding -- 5.4 Graph layout -- 5.4.1 Force directed
graph layout: Elastic Embedding -- 5.4.2 Probabilistic graph layout:
LargeVis -- 5.4.3 Topological method UMAP -- 5.5 Artificial neural
networks -- 5.5.1 Auto-encoders -- 5.5.2 IVIS -- 6 Supervised DR --
6.1 Types of supervision -- 6.1.1 Full supervision -- 6.1.2 Weak
supervision -- 6.1.3 Semi-supervision -- 6.2 Parametric with class
purity -- 6.2.1 Linear Discriminant Analysis -- 6.2.2 Neighbourhood
Component Analysis -- 6.3 Metric learning -- 6.3.1 Mahalanobis
distances -- 6.3.2 Riemannian metric -- 6.3.3 Direct distances
transformation -- 6.3.4 Similarities learning -- 6.3.5 Metric learning
limitations -- 6.4 Class adaptive scale -- 6.5 Classimap -- 6.6 CGNE
-- 6.6.1 ClassNeRV stress -- 6.6.2 Flexibility of the supervision --
6.6.3 Ablation study -- 6.6.4 Isolet 5 case study -- 6.6.5 Robustness
to class misinformation -- 6.6.6 Extension to the type 2 mixture:
ClassJSE -- 6.6.7 Extension to semi-supervision and weak-supervision
-- 6.6.8 Extension to soft labels -- 7 Mapping construction -- 7.1
Optimization -- 7.1.1 Global and local optima -- 7.1.2 Descent
algorithms -- 7.1.3 Initialization -- 7.1.4 Multi-scale optimization --
7.1.5 Force-directed placement interpretation -- 7.2 Acceleration
strategies -- 7.2.1 Attractive forces approximation -- 7.2.2 Binary
search trees -- 7.2.3 Repulsive forces -- 7.2.4 Landmarks
approximation -- 7.3 Out of sample extension -- 7.3.1 Applications --
7.3.2 Parametric case -- 7.3.3 Non-parametric stress with neural
network model -- 7.3.4 Non-parametric case -- 8 Applications -- 8.1
Smart buildings commissioning -- 8.1.1 System and rules -- 8.1.2
Mapping -- 8.2 Photovoltaics -- 8.2.1 I–V curves -- 8.2.2 Comparing
normalized I–V curves -- 8.2.3 Colour description of the chemical
compositions -- 8.3 Batteries -- 8.3.1 Case 1 1 -- 8.3.2 Case 2 2 -- 9
Conclusions -- Nomenclature -- A Some technical results -- A.1
Equivalence between triangle inequality and convexity of balls for -- a
pseudo-norm -- A.2 From Pareto to exponential distribution -- A.3
Spiral and Swiss roll -- B Kullback–Leibler divergence -- B.1
Generalized Kullback–Leibler divergence -- B.1.1 Perplexity with hard
neighbourhoods -- B.2 Link between soft and hard recall and precision
-- Details of calculations -- C.1 General gradient of stress function --

C.2 Neighbourhood embedding -- C.2.1 Supervised neighbourhood embedding (asymmetric case) -- C.2.2 Mixtures -- C.2.3 Belonging rates -- C.2.4 Soft-min arguments -- C.2.5 Scale setting by perplexity -- C.2.6 Force interpretation -- D Spectral projections algebra -- D.1 PCA as matrix factorization and SVD resolution -- D.2 Link with linear projection -- D.3 Sparse expression -- D.4 PCA and centering: from affine to linear -- D.5 Link with covariance and Gram matrices -- D.6 From distances to Gram matrix -- D.6.1 Probabilistic interpretation and maximum likelihood -- D.7 Nyström approximation -- References -- Index 7.

Sommario/riassunto

This book proposes tools for analysis of multidimensional and metric data, by establishing a state-of-the-art of the existing solutions and developing new ones. It mainly focuses on visual exploration of these data by a human analyst, relying on a 2D or 3D scatter plot display obtained through Dimensionality Reduction (DR). Performing diagnosis of an energy system requires identifying relations between observed monitoring variables and the associated internal state of the system. Dimensionality reduction, which allows to represent visually a multidimensional dataset, constitutes a promising tool to help domain experts to analyse these relations. This book reviews existing techniques for visual data exploration and dimensionality reduction, and proposes new solutions to challenges in that field. In order to perform diagnosis of energy systems, domain experts need to establish relations between the possible states of a given system and the measurement of a set of monitoring variables. Classical dimensionality reduction techniques such as tSNE and Isomap are presented, as well as the new unsupervised technique ASKI and the supervised methods ClassNeRV and ClassJSE. A new approach, MING for local map quality evaluation, is also introduced. These methods are then applied to the representation of expert-designed fault indicators for smart-buildings, I-V curves for photovoltaic systems and acoustic signals for Li-ion batteries.
