Test Problems

In this work, we evaluate and compare the proposed methods on benchmark test problems. The data sets are summarized in the following.

A.1 Astronomy

Our experimental study is based on data from the Sloan Digital Sky Survey (SDSS) [2]. In [6], the EFIGI catalog of 4,458 nearby galaxies is presented. We concentrate on a subset of 100 galaxies of types -6 to -4 and 7 to 11. The images are 40 × 40 RGB images. The data has been preprocessed: all entries below a threshold $\psi = 0.1$ have been set to 0. Figure A.1 shows a selection of eight galaxies from the EFIGI data set. The galaxy data set is employed in the experimental analysis of Chapter 5.

![Fig. A.1 Visualization of eight sample galaxies from the EFIGI data set](image)

A.2 Boston

The Boston housing data set stems from 506 census tracts of Boston in 1970. It consists of $N = 506$ patterns with $d = 13$ features (positive real values), e.g., proportion of owner-occupied units built prior to 1940 and weighted distances to five Boston employment centers. The original data has been published by Harrison and Rubinfeld [38]. The data set was taken from the StatLib library, which is maintained at Carnegie Mellon University.
A.3 Digits

The *Digits* data set \[46\] comprises handwritten digits and is often employed as reference problem related to the recognition of handwritten characters and digits. Figure [A.2] shows a collection of images from the *Digits* data set. The collection shows all ten digits, while most experimental analyses concentrate on a label subset.

![Fig. A.2 Visualization of a collection of images from the Digits data set](image)

A.4 NIALM Data

The NIALM data sets have been recorded in collaboration with the OFFIS\[1\], Oldenburg. The *install* data set consists of 120 patterns that have manually been recorded and labeled. The *field study* data set consists of patterns that have been recorded in a household test environment. Table A.1 shows the appliances that are part of the test data sets *install* and of the *field study*.

Table A.1 List of 15 appliances of the *install* and the *field study* data set

<table>
<thead>
<tr>
<th># appliances</th>
<th># appliances</th>
<th># appliances</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 shelf light</td>
<td>6 table light, bedroom</td>
<td>11 ceiling lamp, bathroom</td>
</tr>
<tr>
<td>2 fridge</td>
<td>7 table light, TV</td>
<td>12 ceiling lamp, living room</td>
</tr>
<tr>
<td>3 bedside lamp</td>
<td>8 table light, door</td>
<td>13 ceiling lamp, corridor</td>
</tr>
<tr>
<td>4 desk lamp</td>
<td>9 kettle</td>
<td>14 ceiling lamp, bedroom</td>
</tr>
<tr>
<td>5 TV</td>
<td>10 mirror lamps</td>
<td>15 air conditioning</td>
</tr>
</tbody>
</table>

A.5 ISOMAP-Faces

The ISOMAP-*Faces* data set has been used in the experimental study of the original ISOMAP article \[103\]. It consists of \(N = 698\) images of faces of a statue with varying **poses** and **lights**. The data set can be downloaded from [http://isomap.stanford.edu/](http://isomap.stanford.edu/). Figure [A.3] shows the first 45 images of the data set.

---

\[1\] Oldenburger Forschungs- und Entwicklungsinstitut für Informatik.
Fig. A.3 Visualization of a collection of images from the *Faces* data set

A.6 S-Structure

The 3-dimensional $S$ data set consists of 500 points in the version 3D-S without hole, see Part (a) of Figure A.4. The counterpart 3D-S$_h$ with hole consists of approximately 350 points.

Fig. A.4 3-dimensional $S$: (a) the 3D-S data set and (b) the 3D-S$_h$ data set with hole


88. Pearson, K.: On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2(6), 559–572 (1901)
Index

ε-insensitive loss, 63

ant colony optimization, 75
artificial intelligence, 1 3
astronomy, 2

bagging, 25
balltree, 19
bandwidth matrix, 44
boosting, 25

classification, 4 13
clustering, 5 64 104
CMA-ES, 78
co-ranking matrix, 50 89 97 99 100
covariance matrix, 39 78
cross-validation, 18
curse of dimensionality, 19 83 100
dimensionality reduction, 33 36 39 55
discriminant function, 26
DSRE, 12 14 56 61 69 86
Dunn index, 35
dynamic time warping, 6

EA, 76
eigenvalue, 39
embed-and-repair, 69
embedding, 33 41 58
emergence, 34
empirical risk, 17
ensemble, 68 69

evolutionary algorithm, 76

feature selection, 34
Gaussian distribution, 77 93
gesture recognition, 6

hard margin, 26
Hughes effect, 19 83 100
hybridization, 68 69

imputation, 68
inner product, 26
intrinsic structure, 33
ISOMAP, 40
isotropic Gaussian mutation, 78

k-d tree, 19
K-means, 34 104
K-nearest neighbors, 4 14 25

kernel

bandwidth, 45
function, 44 96
regression, 43
KNN, 4 14 25 56
classification, 14
distance-weighted, 21
ensemble, 25
propagating one, 22
regression, 15
KUNN, 96

latent
representation, 33
Index

sorting, 57
variable, 33
leave-one-out cross-validation, 47
linear
classifier, 26
separation, 26
LLE, 41
local optimum, 36
locally linear embedding, 41
LOO-CV, 47
manifold clustering, 83
manual intelligence, 7
margin, 26
Minkowski, 14 64
missing data, 67
multi-class, 13
mutation, 77
Nadaraya-Watson estimator, 43
nearest neighbor queries, 87
neighborhood function, 37
NIALM, 28
no-free-lunch, 25
nonintrusive appliance load monitoring, 28
novelty detection, 4
offspring, 76
outlier detection, 8
overfitting, 3 16 17 46
p-norm, 14 64
particle swarm optimization, 85
PCA, 29 103
population, 74
principal component analysis, 103
propagating 1-nearest neighbor, 22
PSEA, 86
PSO, 85
quantization error, 50
Rechenberg’s rule, 78
recombination, 76
regression, 15 43
regularized risk, 4
repair-and-embed, 68
robust loss, 63
scalar product, 26
selection, 78
coma, 78
plus, 78
self-organizing map, 6 36 86
semi-supervised learning, 22 68
SL-UNN, 106
slack variable, 27
SOM, 6 86 88
statistical learning theory, 16
stigmergy, 84
submanifold learning, 106
supervised learning, 3
support vector machine, 26 96
swarm intelligence, 84
symbol grounding, 9
time series, 5
topographic error, 50
UKR, 43 46
ULPR, 48
UNN, 55
embed-and-repair, 69
evolutionary, 75
kernel learning, 93
KUNN, 93
noise handling, 63
repair-and-embed, 68
unsupervised kernel regression, 43 46
unsupervised learning, 5
unsupervised nearest neighbors, 55
Vicon, 7
wind energy, 5 118
winner neuron, 36