

Index

A

Abstraction, 3–5
Anomaly detection
 algorithms, 145
 clustering learning problem, 149
 methods, 150
 ML.NET offers, 149, 150
 spot spikes, sales data, 153–157
 statistical problem, 148
 supervised learning problem, 148
Automatic captioning of photographs, 160
Average distance (AD), 110

B

Binary classification, 73, 88
Boston Housing Dataset, 64
Boston housing price prediction
 regression model, 68, 69
Boston housing problem
 ModelInput, 65–67
 ModelOutput, 67
Bounding box, 162, 163, 165
BuildPipeline method, 63

C

Categorical data, 39
Categorical transformations, 41
Centroid, 98–100
 clustering Iris flowers, 102–107

 locations, 107, 108
 model with ground truths, 109, 110

Classifications

 case study, 74–82
 ML.NET trainers, 87
 model evaluation
 confusion matrix, 83–85
 log loss, 86, 87
 setting up options, 89–91
 trainers, 73
 types, 73

Clustering

 Iris flowers, .NET, ml, 102–107
 K-Means (*see* K-Means clustering algorithm)
 learning problem, 149

Collaborative filtering (CF), 131, 133

Collective anomaly, 147

Confusion matrix, 83–85

Content-based filtering (CBF), 131

Contextual anomaly, 147

Convolutional neural net (CNN), 151

Cross-entropy, 86

Customer segmentation, 95

D

DataContext, 6

Data handling

 categories, 45

 cheat sheet, 50

INDEX

Data handling (*cont.*)

- featurization, 48
- missing values, 49, 50
- normalization, 45, 46
- removing, 46, 47

Data loaders, 7, 25–27, 29, 30

Data scrubbing, 44

Data transformation analogy, 7, 37

Data types, 38

Davies-Bouldin index (DBI), 111

E

Emotion analysis, 119

Emotion detection, 119

Euclidean distance, 96–99

F

FastTreeOva algorithm, 80

Featurization, 48

FeaturizeText, 43, 48, 50

Fit method, 5

G

Generic linear regression, 54

Grid, 160, 162

H

Handling missing values, 49, 50

Handling negations, 118

House prices predicting in Boston
suburbs, 64–69

I, J

IDataView, 26, 27, 31, 40, 105

IndicateMissingValues, 63

InputOutputColumnPair, 40

Intersection over union (IOU), 165

K

K-Means clustering algorithm

- centroid, 98–100

- Euclidean distance, 97

- initialization, 101

- mathematics, 95

- notion of proximity, 96

- third axis and beyond, 96

KMeansPlusPlus, 112

L

Lexicons, 114

Linear regression, 4, 55

Loading data, databases, 31

Logarithmic loss, 86

Log loss, 86, 87

M

Matrix factorization

- big matrix, finding, 135

- definition, 132

- let's predict, 135

- Model Builder, 137–144

- modifying hyperparameters, 136

- space advantage, 134

Mean squared error, 70

Miles per gallon (MPG)

- code walkthrough, 63

- dataset, 56

- Model Builder wizard

 - add Machine Learning, 56

 - autogenerated code, 61

- evaluation report, 59
 - execution, 59
 - generated ModelInput.cs, 62
 - training data loaded, 58
 - value prediction, 57
 - Min-max normalization equation, 40
 - MLContext, 108
 - MLContext type, 6, 24
 - MLContext class, 24, 25
 - MLContext instance, 26
 - ML.NET
 - data loading, 7
 - data transformation, 7
 - democratizes, 3
 - encapsulation, 4, 151
 - measuring accuracy, 7
 - pipeline (*see* Pipeline)
 - predict house prices in Boston
 - suburbs, 64
 - prediction, 7
 - problems types, 5
 - “supervised learning” and
 - “unsupervised learning”, 5
 - Model Builder (automatic ML)
 - fly generated UI, 17, 19
 - generated code, 18–20
 - Iris app, creation, 10
 - Iris flower dataset, 8
 - real-world machine learning
 - problem, 8
 - setting up, 13
 - training model, 15
 - training wizard, progress, 14
 - train model interface, 14
 - wizard interface, 10
 - add generated projects to
 - solution, 17
 - loading training data, 11
 - previewing and setup, 12
 - wizard model evaluate interface, 16
 - Model Builder (automatic ML)UCI
 - database, data loading, 8
 - Multiclass classification, 74
 - Multiclass classifiers, 88, 89
- ## N
- Non-maximum
 - Suppression (NMS), 164–166
 - Normalization, 40, 45, 46
 - Normalized mutual
 - information (NMI), 111
 - Normalized root mean square, 71, 72
 - NormalizeMinMax transformations, 40, 69
 - Notion of proximity, 96
 - NormalizeText, 44
 - Numerical data, 38
- ## O
- Object counting, 159
 - Object detection, 159, 168, 169
 - One-hot encoding, 41, 42
 - ONNX Model, 170
 - OnnxObjectDetectionWeb, 169
 - Organized buyers, 93, 94
 - Overfitting, 54
 - Overlapping bounding boxes, 164
 - Oversimplification, 4
- ## P, Q
- Parameters, SRCNN method, 152
 - Performance metrics, 69, 70
 - Pipeline
 - data loaders, 25–27, 29, 30

INDEX

Pipeline (*cont.*)

- loading data from databases, 31
- MLContext, 24, 25
- ModelBuilder, 35
- trainers, 33, 34
- transformers, 31, 32

Point anomaly, 146

Polynomial form, 55

Polynomial regression, 55

Proximity, 1

R

Recommender systems, 131

RegressionMetrics class, 69

Regressions, 4, 5, 53, 56

- description, 54

- equation, 54

- MPG value, predict, 55

- value prediction, 56

Reinforcement learning, 3

ReplaceMissingValues, 63

Root mean square, 71

R-squared, 69, 70

S

Sense of proximity, 1

Sentiment analysis

- generalization, 119–127

- handling negations, 118

- lexicon/word, 114

- polarity, 116, 117

- SentiWordNet dictionary, 115

- unpredictable, 114

SentiWordNet, 114, 118

SquareLossRegression, 136

SRCNN algorithm, 150, 151

Supervised learning problem, 1, 3, 5, 148

T

Test data, 74

TextLoader, 29

Textual data, 38, 43, 44

Trainers, 33, 34

Transformation

- categorical data, 41–43

- numerical data, 39–41

- textual data, 43, 44

TransformCatalog, 7

Transformers, 31, 32

U

Unsupervised learning, 3, 6

V, W, X

Video surveillance (real-time object detection), 159

Y, Z

You Only Look Once (YOLO)

- bounding box, 162, 163

- input dimensions, 169

- input image, 161

- V2 model via ML.NET, 166–168