

Matice v štatistike a dátovej vede

Samuel Rosa

Fakulta matematiky, fyziky a informatiky
Univerzity Komenského v Bratislave

Prvé kapitoly učebníc viacrozmernej štatistiky, strojového učenia, dátovej vedy:

Prvé kapitoly učebníc viacrozmernej štatistiky, strojového učenia, dátovej vedy:

Izenman (2008): Modern Multivariate Statistical Techniques

xii	Contents	
2.2	Examples	18
2.2.1	Example: DNA Microarray Data	18
2.2.2	Example: Mixtures of Polyaromatic Hydrocarbons	19
2.2.3	Example: Face Recognition	22
2.3	Databases	24
2.3.1	Data Types	25
2.3.2	Trends in Data Storage	26
2.3.3	Databases on the Internet	27
2.4	Database Management	29
2.4.1	Elements of Database Systems	29
2.4.2	Structured Query Language (SQL)	30
2.4.3	OLTP Databases	32
2.4.4	Integrating Distributed Databases	32
2.4.5	Data Warehousing	33
2.4.6	Decision Support Systems and OLAP	35
2.4.7	Statistical Packages and DBMSs	36
2.5	Data Quality Problems	36
2.5.1	Data Inconsistencies	37
2.5.2	Outliers	38
2.5.3	Missing Data	39
2.5.4	More Variables than Observations	40
2.6	The Curse of Dimensionality	41
	Bibliographical Notes	42
	Exercises	43
3	Random Vectors and Matrices	45
3.1	Introduction	45
3.2	Vectors and Matrices	45
3.2.1	Notation	45
3.2.2	Basic Matrix Operations	46
3.2.3	Vectoring and Kronecker Products	47
3.2.4	Eigenanalysis for Square Matrices	48
3.2.5	Functions of Matrices	49
3.2.6	Singular-Value Decomposition	50
3.2.7	Generalized Inverses	50
3.2.8	Matrix Norms	51

Contents

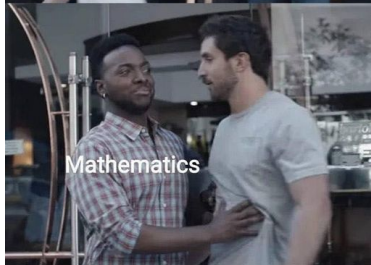
Website	vii
Acknowledgments	viii
Notation	xi
1 Introduction	1
1.1 Who Should Read This Book?	8
1.2 Historical Trends in Deep Learning	11
1 Applied Math and Machine Learning Basics	29
2 Linear Algebra	31
2.1 Scalars, Vectors, Matrices and Tensors	31
2.2 Multiplying Matrices and Vectors	34
2.3 Identity and Inverse Matrices	36
2.4 Linear Dependence and Span	37
2.5 Norms	39
2.6 Special Kinds of Matrices and Vectors	40
2.7 Eigendecomposition	42
2.8 Singular Value Decomposition	44
2.9 The Moore-Penrose Pseudoinverse	45
2.10 The Trace Operator	46
2.11 The Determinant	47
2.12 Example: Principal Components Analysis	48
3 Probability and Information Theory	53
3.1 Why Probability?	54

Grus (2015): Data Science from Scratch

Truthiness	25
The Not-So-Basics	26
Sorting	27
List Comprehensions	27
Generators and Iterators	28
Randomness	29
Regular Expressions	30
Object-Oriented Programming	30
Functional Tools	31
enumerate	32
zip and Argument Unpacking	33
args and kwargs	34
Welcome to DataSciencester!	35
For Further Exploration	35
3. Visualizing Data	37
matplotlib	37
Bar Charts	39
Line Charts	43
Scatterplots	44
For Further Exploration	47
4. Linear Algebra	49
Vectors	49
Matrices	53
For Further Exploration	55
5. Statistics	57
Describing a Single Set of Data	57
Central Tendencies	59
Dispersion	61
Correlation	62
Simpson's Paradox	65
Some Other Correlational Caveats	66
Correlation and Causation	67
For Further Exploration	68
6. Probability	69
Dependence and Independence	69
Conditional Probability	70
Bayes's Theorem	72
Random Variables	73

Matice dôležité v:

- metódy štatistiky, strojového učenia
- optimalizácia, siete/grafy, ...
- množstvo predmetov DAV, PMA, a magisterských
- vid' http://www.iam.fmph.uniba.sk/ospm/Rosa/MP/MP_motivacia.pdf





Me when I realise I need to learn linear algebra for machine learning

