Foundations of Data Science Syllabus

Name	Weekly amount	Requirement	Credit	Semester
IPM-18fatFDSE	2 lectures	Exam	2	1st
IPM-18fatFDSG	2 practice sessions	Practice grade	2	

1.	Program title	Computer Science MSc
2.	Department	Numerical Analysis
3.	Lecturer	Lajos Lóczi, associate professor

4. Topics

Vector spaces. Subspaces, linear dependence and independence, bases, dimension, codimension, hyperplane.
Linear maps between vector spaces. Null space, image. The rank-nullity theorem. Regular and singular linear maps. Matrices as linear maps. The determinant and the trace. Examples: rotations, reflections, projections, differential and integral operators, shift operators in finite and infinite dimensions.
The rank. Eigenvalues, eigenvectors, eigenspace. Spectrum, characteristic polynomial. Algebraic and geometric multiplicities. Similarity. The Cayley–Hamilton theorem.
Matrix functions, the definition via convergent Taylor series, or via interpolation.
Multilinear maps. Symmetric and antisymmetric maps. The geometric meaning of the determinant, and some applications.
Matrix decompositions, diagonalizability, the Jordan canonical form.

Inner-product spaces. Orthogonality, angle. The Pythagorean theorem. The Cauchy–Schwarz–Bunyakovsky inequality. The Gram–Schmidt process.

Normed vector spaces. Equivalent norms. Examples, induced norms. The operator norm.

Metric spaces. Examples, the induced metric. Cauchy sequences. Complete metric spaces.

Banach spaces and Hilbert spaces. Examples. Fourier series in Hilbert spaces.

Overdetermined linear systems, the method of least squares. Projections. Generalized inverses: the Moore–Penrose pseudoinverse. Definiteness of matrices. The singular-value decomposition (SVD), and PCA.

Probability spaces. Measures. Discrete probability spaces.
Conditional probability. Independent events, mutually exclusive events. Law of total probability. Bayes' theorem. Discrete and continuous random variables. Probability mass function, probability density function, cumulative distribution function. Expected value, variance, and standard deviation of a random variable. Markov's inequality.

	Univariate descriptive **statistics**: central tendency/location (mean, median, mode) and dispersion (range, variance, standard deviation, quantiles, quartiles). Bivariate descriptive statistics: quantitative measures of dependence (covariance, Pearson's correlation, Spearman's correlation) Data visualization: histogram, scatter plot, contingency table. **Entropy** of a random variable.	
5.	Recommended books • Gilbert Strang, *Introduction to Linear Algebra*, 5th Edition, 2016, Wellesley-Cambridge Press, ISBN: 978-09802327-7-6 • K. W. Gruenberg, A. J. Weir, *Linear Geometry*, 2nd Edition, 1977, Springer, ISBN-13: 978-0387902272 • William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1-2, 3rd Edition, 2008, John Wiley & Sons Inc., ISBN-13: 978-8126518050 • John A, Rice, *Mathematical Statistics and Data Analysis*, 3rd Edition, 2010, Cengage Learning, ISBN-13: 978-8131519547 • Thomas M. Cover, Joy A. Thomas, *Elements of Information Theory*, 2nd Edition, 2006, Wiley Series in Telecommunications and Signal Processing, ISBN-13: 978-0471241959	
6.	Course requirements	
	Practice grade	1. Attendance at the practice sessions is mandatory. One can miss at most 3 practice sessions (online or traditional) during the semester. 2. There will be two midterms, M1 (around Week 7) and M2 (near the semester end), consisting of problem solving. If M1 and M2 are both successful (both of them at least 40%), one obtains their grade by averaging the percentages as follows: average 40–54% = Grade 2; 55–69% = Grade 3; 70–84% = Grade 4; 85–100% = Grade 5. If one fails exactly one midterm, they can retake it (R) at the end of the semester. If R is successful, then the final grade is 2 or 3. If one fails both M1 and M2, or R, their grade is 1. In this case, they retake the whole material in a single test. If this is successful, their second grade is 2, otherwise 1.
	Exam	1. After obtaining a successful practice grade (that is, Grade 2–5), one must take an exam. 2. The exam type is oral. The questions will cover definitions, relations among the various concepts, theorems, and some proofs.
7.	Office hours for the **lecturer**	To make an appointment, please contact the lecturer in email (LLoczi@inf.elte.hu).