(*faculty stamp*)

COURSE DESCRIPTION

1. Course title: STATISTICS FOR DATA SCIENCE, Models with hidden data	**2. Course code SFDS_MWHD**
3. Validity of course description: 2018/2019	
4. Level of studies: MSc programme	
5. Mode of studies: intramural studies	
6. Field of study: CONTROL, ELECTRONIC AND INFORMATION ENGINEERING (MACRO)	(FACULTY SYMBOL) RAU-2
7. Profile of studies: ACADEMIC	
8. Programme: DATA SCIENCE	
9. Semester: 2	
10. Faculty teaching the course: Faculty of Automatic Control, Electronics and Computer Science	
11. Course instructor: Prof. dr hab. inż. Andrzej Polański	
12. Course classification: common courses	
13. Course status: compulsory ~~/elective~~	
14. Language of instruction: English	
15. Pre-requisite qualifications: Algebra and analytic geometry, Calculus and differential equations, Physics, Computer programming, Optimization methods, Numerical methods, Statistics and probability theory, Algorithms and data structures.	
16. Course objectives: The aim of the course is making students familiar with issues related to statistical models with hidden variables.	
17. Description of learning outcomes:	

Nr	Learning outcomes description	Method of assessment	Teaching methods	Learning outcomes reference code
1.	Student understands the notion of latent variable in the stochastic model.	Credit	Lecture	K2A_W01, K2A_W02
2.	Student understands the EM algorithm for estimation of parameters of models with hidden data.	Credit	Lecture	K2A_W06, K2A_W07
3.	Student is able to elaborate the algorithm and software for parameter estimation with the use of EM iterations.	Laboratory tasks	Laboratory	K2A_U06, K2A_U07, K2A_K01
4.	Student is able to elaborate software for sematic analyses with latent variables.	Laboratory tasks	Laboratory	K2A_U03, K2A_U09, K2A_K02
5.	Student is able to use existing software and to elaborate algorithms for estimation of parameters of mixtures of hidden Markov models.	Laboratory tasks	Laboratory	K2A_U03, K2A_K01
6.				
7.				

8.				
9.				

18. Teaching modes and hours

Lecture 15 / ~~BA /MA Seminar / Class / Project~~ / Laboratory 15

19. Syllabus description:

Lecture:

1. Introductory issues. Notions of missing, latent, hidden variables in data analyses. Examples of problems involving latent variables in science, engineering, optimization, parameter estimation.
2. Mixtures of Gaussian distributions. Parameters. Gaussian components. Component parameters. Component index as latent variable. Bayesian formula for computing conditional distribution of latent variables. Intuitive derivation of estimates of component parameters.
3. General expectation maximization algorithm for iterative likelihood maximization. Complete observations, incomplete observations, latent observations. Conditional distribution of hidden variables given data and parameters guess. Integral formula for log likelihood. Expectation step. Maximization step. Jensen's inequality. Properties of the EM algorithm.
4. EM algorithm for mixtures of distributions. EM algorithm for censored data.
5. Probabilistic latent semantic analysis. Co-occurrence tables. Aspect model. Latent semantic analysis. Application of the EM algorithm.
6. Markov models with latent states. Hidden Markov models. Baum – Welch algorithm.
7. Mixtures of Markov models and hidden Markov models.

Laboratory:

1. Mixtures of normal distributions
2. Probabilistic latent semantic analysis
3. Mixtures of Markov models and hidden Markov models.

20. Examination: semester NO

21. Primary sources:

G.J. McLachlan, T Krishnan, (2008), The EM Algorithm and Extensions, Wiley
G.J. McLachlan, D. Peel, (2000), Finite Mixture Models, Wiley

22. Secondary sources:

C.M. Bishop, (2006), Pattern Recognition and Machine Learning, Springer

23. Total workload required to achieve learning outcomes

Lp.	Teaching mode :	Contact hours / Student workload hours
1	Lecture	15/15
2	Classes	/
3	Laboratory	15/15
4	Project	/
5	BA/ MA Seminar	/
6	Other	/
	Total number of hours	30/30

24. Total hours: 60

25. Number of ECTS credits: 2

26. Number of ECTS credits allocated for contact hours: 1

27. Number of ECTS credits allocated for in-practice hours (laboratory classes, projects): 1

26. Comments:

Approved:

……
(date, Instructor's signature)

……
(*date , the Director of the Faculty Unit signature)*