



# Pattern Recognition Methods and Introduction to Machine Learning

## Course description sheet

### Basic information

|                                                                                                                                                                                                                                                                                                  |                                                                                                                                             |                                                                                                                                                                                                                                                                                                  |                                    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|
| <b>Field of study</b><br>AGH UST International Courses<br><b>Major</b><br>All<br><b>Organisational unit</b><br>AGH University Database of Electives<br><b>Study level</b><br>University database of electives<br><b>Form of study</b><br>Full-time studies<br><b>Profile</b><br>General academic |                                                                                                                                             | <b>Didactic cycle</b><br>2024/2025<br><b>Course code</b><br>UBPOJOS.A200000.04820.24<br><b>Lecture languages</b><br>English<br><b>Mandatoriness</b><br>Obligatory<br><b>Block</b><br>General Modules<br><b>Course related to scientific research</b><br>No<br><b>USOS code</b><br>240-INT-xS-125 |                                    |
| <b>Course coordinator</b>                                                                                                                                                                                                                                                                        | Witold Dzwinel                                                                                                                              |                                                                                                                                                                                                                                                                                                  |                                    |
| <b>Lecturer</b>                                                                                                                                                                                                                                                                                  | Witold Dzwinel                                                                                                                              |                                                                                                                                                                                                                                                                                                  |                                    |
| <b>Period</b><br>Summer semester                                                                                                                                                                                                                                                                 | <b>Method of verification of the learning outcomes</b><br>Exam<br><br><b>Activities and hours</b><br>Lectures: 14<br>Laboratory classes: 14 |                                                                                                                                                                                                                                                                                                  | <b>Number of ECTS credits</b><br>4 |

## Goals

|    |                                                                                                                                     |
|----|-------------------------------------------------------------------------------------------------------------------------------------|
| C1 | To familiarize students with the role of machine learning in modern technology and its role in artificial intelligence              |
| C2 | To make students acquainted with machine learning terminology.                                                                      |
| C3 | Familiarize them with basic ML tasks: feature engineering, classification, regression, clustering algorithms and validation methods |
| C4 | To make students acquainted with the principles of developing simple ML systems.                                                    |

## Course's learning outcomes

| Code                                              | Outcomes in terms of                                                                                                                                    | Learning outcomes prescribed to a field of study | Methods of verification                                                                                                                   |
|---------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Knowledge - Student knows and understands:</b> |                                                                                                                                                         |                                                  |                                                                                                                                           |
| W1                                                | Student knows and understands methods of data normalization and calibration                                                                             |                                                  | Activity during classes, Participation in a discussion, Execution of a project, Execution of laboratory classes, Examination, Oral answer |
| W2                                                | Students have the knowledge allowing them to choose a proper data representation and use feature engineering to avoid "curse of dimensionality" problem |                                                  | Activity during classes, Participation in a discussion, Examination, Oral answer                                                          |
| W3                                                | Students have an extended expertise in classification, regression or clustering algorithms, validation and data visualization methods                   |                                                  | Activity during classes, Execution of laboratory classes, Examination, Engineering project, Completion of laboratory classes              |
| <b>Skills - Student can:</b>                      |                                                                                                                                                         |                                                  |                                                                                                                                           |
| U1                                                | Students can develop an ML data analysis system based on classification, regression or clustering and validate the results.                             |                                                  | Engineering project, Completion of laboratory classes                                                                                     |
| U2                                                | The ability of using ML Python libraries.                                                                                                               |                                                  | Execution of laboratory classes, Engineering project, Completion of laboratory classes                                                    |
| <b>Social competences - Student is ready to:</b>  |                                                                                                                                                         |                                                  |                                                                                                                                           |
| K1                                                | The improvement of skills concerning public speaking.                                                                                                   |                                                  | Participation in a discussion                                                                                                             |
| K2                                                | The development of public relations through discussions with opponents.                                                                                 |                                                  | Activity during classes, Participation in a discussion                                                                                    |
| K3                                                | Student can present his/her results to the audience.                                                                                                    |                                                  | Activity during classes, Participation in a discussion                                                                                    |

## Program content ensuring the achievement of the learning outcomes prescribed to the module

Students will be provided with basics in machine learning and data analysis. They will understand the methodological construction of ML systems and to start more advanced courses.

### Student workload

| Activity form                                       | Average amount of hours* needed to complete each activity form |
|-----------------------------------------------------|----------------------------------------------------------------|
| Lectures                                            | 14                                                             |
| Laboratory classes                                  | 14                                                             |
| Preparation for classes                             | 15                                                             |
| Realization of independently performed tasks        | 15                                                             |
| Examination or final test/colloquium                | 5                                                              |
| Contact hours                                       | 28                                                             |
| Preparation of project, presentation, essay, report | 15                                                             |
| <b>Student workload</b>                             | <b>Hours</b><br>106                                            |
| <b>Workload involving teacher</b>                   | <b>Hours</b><br>28                                             |

\* hour means 45 minutes

### Program content

| No. | Program content                                                                                                                                                                                                                                                                                                                                                                                                                                           | Course's learning outcomes | Activities                   |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------|------------------------------|
| 1.  | Introduction<br><br>1. Artificial Intelligence methods in the modern world. AI taxonomy.<br>2. Who is a data scientist?<br>3. What is Machine Learning and why it is crucial for data analysis.<br>4. Five general principles data scientists should know Benfords Law, Law of Large Numbers, Zip's law, Bonferroni's Principle, Averages definitions.<br>5. ML in a nutshell: feature engineering, regression - classification - clustering, validation. | W1, U2, K2, K3             | Lectures, Laboratory classes |

| No. | Program content                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Course's learning outcomes | Activities                   |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------|------------------------------|
| 2.  | <p>Data representation</p> <ol style="list-style-type: none"> <li>1. Feature vector and feature space.</li> <li>2. Features' normalization.</li> <li>3. The properties of the high-dimensional spaces.</li> <li>4. Curse of dimensionality. Strange properties of N-D spaces.</li> <li>5. Feature engineering and data embedding --&gt; how to diminish the dimensionality (basics)</li> <li>6. Non-structured data. Sophisticated embedding.</li> </ol>                                                                                                 | W2, U2, K2                 | Lectures, Laboratory classes |
| 3.  | <p>Data/Model representation</p> <ol style="list-style-type: none"> <li>1. ML tasks: Classification - Regression and Clustering</li> <li>2. Number of linear dichotomies. Underfit.</li> <li>3. Vapnik-Chervonienkis dimension.</li> <li>4. Distances and proximity measures.</li> <li>5. Vector-Vector, Vector- Set, Set-Set, Pdf - pdf: Minkowsky distances, Mahalanobis, Proximity measures, Kullback-Leibler divergence, Wasserstein metrics, optimal transport theory.</li> <li>6. Manifold learning irt distance - feature engineering.</li> </ol> | W2, U2, K2                 | Lectures, Laboratory classes |
| 4.  | <p>Model representation</p> <ol style="list-style-type: none"> <li>1. Classification/regression - performance measure - loss functions</li> <li>2. Optimizers (gradient descent, stochastic gradient descent, minibatch SGD, momentum, Nesterov, Adam, Adagrad, Adadelata etc.)</li> <li>3. Training, validation, testing: terminology, one hit, cross-validation, ROC, AUC, Precision/Recall, F1.</li> </ol>                                                                                                                                            | W1, W2, U2, K2, K3         | Lectures, Laboratory classes |
| 5.  | <p>Feature engineering</p> <ol style="list-style-type: none"> <li>1. Sparse coding hipotesis.</li> <li>2. Feature selection algorithms (filters, wrappers)</li> <li>3. Feature extraction methods: Principal component analysis (PCA), Linear Discriminant Analysis (LDA), Fischer algorithm</li> <li>4. Autoencoders.</li> <li>5. Random Projection and Johnson-Lindenstruss theorem.</li> <li>6. Non-linear methods: Multidimensional scaling (MDS), Isomap, t-SNE.</li> </ol>                                                                         | W1, W2, U2, K2, K3         | Lectures, Laboratory classes |
| 6.  | <p>Simple classifiers</p> <ol style="list-style-type: none"> <li>1. Minimum distance (maximum discriminant) classifiers</li> <li>2. Bayesian rule and clussifiers - the best classifier.</li> <li>3. Naive Bayes and k-NN classifier.</li> <li>4. Kernel k-NN and the methods of finding nearest neighbors (kd-tree, Fridmann algorithm)</li> </ol>                                                                                                                                                                                                      | W2, U1, U2, K2, K3         | Lectures, Laboratory classes |

| No. | Program content                                                                                                                                                                                                                                                                                                                                                                                                                                  | Course's learning outcomes | Activities                   |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------|------------------------------|
| 7.  | Advanced classifiers<br>1. Linear classifiers (perceptron, Fischer, linear SVM)<br>2. Support Vector Machine classifier (regression).<br>3. Ensemble classifiers.<br>4. Neural Networks (MLP, DNNs, Hopfield, RBM and DBNs)<br>5. Bagging and Boosting.<br>6. Imbalanced classes.                                                                                                                                                                | W3, U1, U2, K1, K3         | Lectures, Laboratory classes |
| 8.  | Clustering<br>1. Expectation maximization algorithm (EM).<br>Generative models (DBN)<br>2. Clustering algorithms - taxonomy.<br>3. Hierarchical algorithms (Ward, Jarvis-Patrick, mutual nearest neighborhood others)<br>4. Partitioning clustering: k-Means, kernel k-Means, fuzzy-k-Means<br>5. Modern clustering schemes: DBSCAN, Chameleon, SNN, CURE, Affinity propagation.<br>6. Clustering validation: purity, entropy, Silhouette plots. | W2, W3, U1, U2, K1, K2     | Lectures, Laboratory classes |
| 9.  | Pattern recognition - basics<br>1. Data acquisition and preparation. Machine Learning vs Pattern Recognition<br>2. Convolution principles, filters<br>3. Main transforms: Fourier, wavelets, Karhunen-Loeve, Hough<br>4. Introduction to Convolutional Neural Networks                                                                                                                                                                           | W1, W3, U1, U2, K2         | Lectures                     |

### Extended information/Additional elements

#### Teaching methods and techniques :

Lectures, Discussion, E-learning

| Activities   | Methods of verification                                                                                                                        | Credit conditions |
|--------------|------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|
| Lectures     | Activity during classes, Participation in a discussion, Execution of a project, Execution of laboratory classes, Examination, Oral answer      |                   |
| Lab. classes | Activity during classes, Participation in a discussion, Execution of laboratory classes, Engineering project, Completion of laboratory classes |                   |

#### Method of determining the final grade

The final grade is calculated as the weighted sum of the grades from oral (or test) exam and the lab activities. The weights are the same and equal to 0.5. The students with excellent (5.0) grade from lab exercises are exempt from oral exam.

## **Prerequisites and additional requirements**

The passed courses in Calculus, Algebra and Numerical Methods. Good knowledge of Python.

### **Rules of participation in given classes, indicating whether student presence at the lecture is obligatory**

Lectures: Students participate in the classes by learning the subsequent contents according to the course syllabus. Students should systematically ask questions and clarify doubts. Audiovisual recording of the lecture requires the consent of the instructor. Laboratory classes: Students perform laboratory exercises following the materials provided by the instructor. The student is required to prepare in the subject of the exercise, which can be verified by a colloquium in oral or written form. Credit for classes is given based on a presentation of solutions to problems posed. The pass mark for the module is given after the successful completion of all laboratory classes.

## **Literature**

### **Obligatory**

1. Bishop, C. M. (2006). Machine learning and pattern recognition. Information Science and Statistics. Springer, Heidelberg.
2. Grossman R., L., Karnath, Ch, Kegelmeyer, P., Kumar, V., Namburu, R.,R., Data Mining for Scientific and Engineering Applications, Kluwer Academic Publisher, 2001
3. R.Duda, P.Hart, D. Stork, Patterns Classification, Wiley Interscience, 2001.

### **Optional**

1. Theodoris S and Koutroumbas K, Pattern Recognition, Academic Press, San Diego, London, Boston, 1998.
2. Mitra, S. and Acharya T., Data Mining: Multimedia, Soft Computing and Bioinformatics, 424pp. J. Wiley, 2003.