

Nazwa pola Field name	Komentarz <sup>1</sup> Comment
Tytuł Course title:	Uczenie statystyczne Statistical machine learning
Jednostka Department:	Wydział Matematyki, Informatyki i Mechaniki UW Faculty of Mathematics, Informatics and Mechanics
Grupy Course groups:	przedmioty obowiązkowe dla pierwszego roku studiów magisterskich uczenie maszynowe Obligatory courses for 1st grade 2nd stage Machine Learning
Kod wg USOS USOS Course ID:	1000-317bSML (klon 1000-1M18SUM)
Kod wg Sokratesa Erasmus code / ISCED:	11.304
Rodzaj zajęć Type of class:	Wykład 30 godzin, Laboratorium 30 godzin Lectures 30h, Laboratories 30h
Opis Short description:	<p>Celem przedmiotu jest wprowadzenie podstawowych pojęć i narzędzi statystycznych używanych w uczeniu maszynowym takich jak regresja liniowa, logistyczna i wielowymiarowa, klasyfikatory, metody redukcji wymiaru, metody bayesowskie.</p> <p>The goal of the course is to introduce fundamental notions and statistical tools used in machine learning such as linear, logistic and multivariate regression, classifiers, dimension reduction methods, bayesian methods.</p>
Program Programme content/Full description	<p>Szczegółowy program</p> <ol style="list-style-type: none"> <li>Statystyka eksploracyjna (2-3 zajęcia) <ol style="list-style-type: none"> <li>Basic summary statistic (wartość średnia, mediana, wariancja próbki, etc.)</li> <li>Wizualizacja danych (histogram, wykres pudełkowy, estymator jądrowy gęstości)</li> <li>Analiza głównych składowych</li> <li>Klasteryzacja, klasteryzacja hierarchiczna, k-średnie, k-medoidy</li> </ol> </li> <li>Teoria statystyki (4-5 zajęć) <ol style="list-style-type: none"> <li>Podstawowe definicje (modele statystyczne, statystyki, wiarygodność, etc.)</li> <li>Teoria szacowania (estymatory największej wiarygodności, efektywność, błąd średniokwadratowy, kompromis między obciążeniem a wariancją, przedziały ufności)</li> <li>Testowanie hipotez statystycznych (błędy typu I i II, moc testu, poziom istotności, wartość p)</li> <li>Problemy związane z wartością p (efekt rozmiaru, testowanie wielu hipotez)</li> <li>Wnioskowanie bayesowskie (rozkład a priori, rozkład a posteriori, ryzyko bayesowskie i estymator bayesowski, przedziały wiarygodności)</li> <li>Odległość między miarami prawdopodobieństwa (dywergencja Kullbacka-Leiblera, miara maksymalnej różnicy, etc.)</li> </ol> </li> <li>Proste modele regresyjne i klasyfikacyjne (3-4 zajęć) <ol style="list-style-type: none"> <li>Regresja liniowa</li> <li>Klasyfikacja. Regresja logistyczna, LDA, QDA</li> <li>Walidacja krzyżowa i bootstrap</li> <li>Wybór modelu i regularyzacja, regresja grzbietowa, metoda Lasso, algorytm forward-backward</li> </ol> </li> <li>Modele zaawansowane (3 zajęcia) <ol style="list-style-type: none"> <li>Modele drzewiaste, bagging, lasy losowe, boosting</li> </ol> </li> </ol>

<sup>1</sup> Przedstawiona jest wstępna wersja sylabusu./ This is a preliminary version of the course outline.

	<p>b. Maszyna wektorów nośnych</p> <p>c. Modele nieliniowe: funkcje sklepane, uogólnione modele addytywne</p> <p>The detailed program</p> <p>5. Explanatory Statistics (2-3 lessons)</p> <p>a. Basic summary statistic (sample mean, median, sample variance, etc.)</p> <p>b. Visualization of data (histogram, box-plots, kernel density estimation)</p> <p>c. Principal component analysis</p> <p>d. Clusterization, hierarchical clustering, k-means, k-medoids</p> <p>6. Statistical theory (4-5 lessons)</p> <p>a. Basic definitions (statistical models, statistics, likelihood, etc.)</p> <p>b. Estimation theory (maximum likelihood estimators, efficiency, mean square error, bias vs variance trade-off, confidence intervals)</p> <p>c. Statistical hypothesis testing (type I and type II errors, power of a test, significance, p-value)</p> <p>d. Problems with p-values (effect size, multiple hypotheses testing)</p> <p>e. Bayesian inference (prior distribution, posterior distribution, Bayesian risk and Bayesian estimator, credible intervals)</p> <p>f. Distance between probability measures (Kullback-Leibler divergence, total variation distance, etc.)</p> <p>7. Simple regression and classification models (3-4 lessons)</p> <p>a. Linear regression</p> <p>b. Classification. Logistic regression, LDA, QDA</p> <p>c. Cross-validation and bootstrap</p> <p>d. Model selection and regularization. Lasso, ridge regression, forward-backwards procedure</p> <p>8. Advanced models (3 lessons)</p> <p>a. Tree-like models, bagging, random forests, boosting</p> <p>b. Support vector machines</p> <p>c. Non-linear models: splines, generalized additive models</p>
<p>Efekty uczenia</p> <p>Learning outcomes</p>	<p>Wiedza: student zna i rozumie</p> <ul style="list-style-type: none"> <li>● w pogłębionym stopniu - podstawową wiedzę z działów matematyki niezbędnych do studiowania uczenia maszynowego (rachunek prawdopodobieństwa, statystyka, analiza wielowymiarowa, i algebra liniowa) [K_W05];</li> <li>● w uporządkowany i podbudowany teoretycznie sposób wiedzę w zakresie podstawowych narzędzi statystycznych stosowanych w modelowaniu i analizie danych [K_W07].</li> </ul> <p>Umiejętności: student potrafi</p> <ul style="list-style-type: none"> <li>● konstruować rozumowania matematyczne [K_U06];</li> <li>● wyrażać problemy w języku matematyki [K_U07];</li> <li>● stosować techniki nowoczesnej statystycznej analizy danych [K_U10].</li> </ul> <p>Kompetencje społeczne: student jest gotów do</p> <ul style="list-style-type: none"> <li>● krytycznej oceny posiadanej wiedzy i odbieranych treści [K_K01];</li> <li>● uznawania znaczenia wiedzy w rozwiązywaniu problemów poznawczych i praktycznych oraz zasięgania opinii ekspertów w przypadku trudności z samodzielnym rozwiązaniem problemu [K_K02];</li> <li>● myślenia i działania w sposób przedsiębiorczy [K_K03].</li> </ul> <p>Knowledge: the student</p> <ul style="list-style-type: none"> <li>● has in-depth understanding of the branches of mathematics necessary to study machine learning (probability theory, statistics, multivariable calculus, and linear algebra) [K_W05];</li> <li>● has based in theory and well organized knowledge of fundamental</li> </ul>

	<p>techniques of statistics used in modeling and data analysis [K_W07].</p> <p>Abilities: the student is able to</p> <ul style="list-style-type: none"> <li>● construct mathematical reasoning [K_U06];</li> <li>● express problems in the language of mathematics [K_U07];</li> <li>● apply techniques of modern statistical data analysis [K_U10].</li> </ul> <p>Social competences: the student is ready to</p> <ul style="list-style-type: none"> <li>● critically evaluate acquired knowledge and information [K_K01];</li> <li>● recognize the significance of knowledge in solving cognitive and practical problems and the importance of consulting experts when difficulties arise in finding a self-devised solution [K_K02];</li> <li>● think and act in an entrepreneurial way [K_K03].</li> </ul>
Proponowane źródła Bibliography:	<p>1. Trevor Hastie, Robert Tibshirani, Jerome H., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, Berlin</p> <p>2. Andrew Ng, <i>Machine Learning Yearning</i>, <a href="https://www.deeplearning.ai/machine-learning-yearning/">https://www.deeplearning.ai/machine-learning-yearning/</a></p>
Kryteria oceniania Assessment criteria:	<p>Test końcowy oraz zadanie programistyczne na ocenę</p> <p>Final test and programming assignment with grades</p>
Wymagania Requirements:	<p>Brak</p> <p>None</p>
Założenia Assumptions	<p>Brak</p> <p>None</p>
Forma zaliczenia Assessment of the learning outcome	<p>Egzamin i zadanie programistyczne na ocenę</p> <p>Exam and graded programming assignment</p>
<i>Teraz informacje o charakterze zmiennym (pole edycji) Variable information (editable fields)</i>	
Cykl dydaktyczny Time span/Study term:	<p>Semestr zimowy 2021/2022</p> <p>Winter semester 2021/2022</p>
Prowadzący Coordinators:	<p>Wykłady ..., Laboratorium ...</p> <p>Lectures: ... Tutorials: ...</p>