Національний технічний університет України
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

ММСА

Кафедра математичних
методів системного аналізу

Інтелектуальний аналіз великих сховищ даних (Big data analytics)

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	***Другий (магістерський)***
Галузь знань	*12 Інформаційні технології*
Спеціальність	*122 «Комп'ютерні науки»*
Освітня програма	*«Системи і методи штучного інтелекту»*
Статус дисципліни	*Вибіркова*
Форма навчання	*очна(денна)*
Рік підготовки, семестр	*1 курс, весняний семестр*
Обсяг дисципліни	*5 кредитів (150 годин), 36 год.лекції, 36 год.лабораторні роботи, 96 год.СРС*
Семестровий контроль/ контрольні заходи	*Екзамен / МКР, письмово*
Розклад занять	*Щотижня лекція і практикум, щотижня заняття з прийому завдань СРС, https://schedule.kpi.ua/*
Мова викладання	*Українська*
Інформація про керівника курсу / викладачів	Лектор: *професор, д.т.н., доц. Недашківська Надія Іванівна, n.nedashkivska@gmail.com* Практичні *заняття*: *професор, д.т.н., доц. Недашківська Надія Іванівна, n.nedashkivska@gmail.com*
Розміщення курсу	Платформа дистанційного навчання «Сікорський», Googleclassroom, код курсу **mqj7y7u**

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Метою кредитного модуля є формування у студентів здатностей:

- *застосовувати сучасні моделі і алгоритми інтелектуального аналізу даних і машинного навчання;*
- *виконувати попередню обробку даних і побудову навчальних наборів;*
- *обирати найбільш важливі ознаки (feature selection), знижувати розмірність даних, отримати суттєві ознаки (feature extraction);*
- *будувати моделі кластеризації даних різної форми, навчати ці моделі, оцінювати якість їх роботи, використовуючи програмне забезпечення python;*
- *прогнозувати споживчий попит на основі наборів даних транзакцій алгоритмами асоціативного аналізу (association mining) та аналізу шаблонів послідовностей (sequential pattern mining), оцінювати якість побудованих асоціативних правил, шаблонів послідовностей і прогнозів на їх основі;*

- *виконувати реконструкцію і породження зображень на основі глибоких моделей автокодувальників; здійснювати попереднє навчання класифікаторів з використанням глибоких автокодувальників;*
- *породжувати нові зображення моделями і методами генеративних змагальних мереж GAN, DCGAN та ін.*
- *будувати моделі рекомендаційних систем різними методами.*

Після засвоєння кредитного модуля мають продемонструвати такі результати навчання:

компетентності:

здатність застосовувати знання в практичних ситуаціях, здатність абстрактно мислити, застосовувати методи аналізу і синтезу, здатність знати та розуміти предметну область і професійну діяльність, здатність до пошуку, оброблення та аналізу інформації з різних джерел, здатність до адаптації та дії в новій ситуації, здатність забезпечувати та оцінювати якість виконуваних робіт,

здатність використовувати системний аналіз в якості сучасної міждисциплінарної методології, заснованої на прикладах математичних методів та сучасних інформаційних технологіях, і орієнтована на вирішення задач аналізу і синтезу технічних, економічних, соціальних, екологічних та інших складних систем,

здатність будувати математично коректні моделі статичних та динамічних процесів і систем із зосередженими та розподіленими параметрами із врахуванням невизначеності зовнішніх та внутрішніх факторів,

здатність до комп'ютерної реалізації математичних моделей реальних систем і процесів; проектувати, застосовувати і супроводжувати програмні засоби моделювання, прийняття рішень, обробки інформації, інтелектуального аналізу даних,

здатність розробляти експериментальні та спостережувальні дослідження і аналізувати дані, отримані в них,

застосовувати методи і засоби роботи з даними і знаннями, методи математичного моделювання, технології системного і статичного аналізу,

проектувати, реалізовувати, тестувати, впроваджувати, супроводжувати, експлуатувати програмні засоби роботи з даними і знаннями в комп'ютерних системах і мережах,

розуміти і застосовувати на практиці методи статичного моделювання і прогнозування, оцінювати вихідні дані;

знання:

методів попередньої обробки даних і побудови навчальних наборів, одновимірних (Univariate Feature Imputation) і багатовимірних алгоритмів (Multivariate Feature Imputation) заповнення відсутніх значень шляхом інтерполяції;

підходів до оцінки важливості ознак з точки зору наявності вчителя, різних стратегій вибору, та з точки зору даних, метод вибору ознак засобами L1-регуляризації, алгоритми послідовного вибору ознак (Sequential Feature Selection), послідовного зворотного вибору (Sequential Backward Selection), рекурсивного виключення ознак (Recursive Feature Elimination), методи вибору ознак для традиційних даних на основі подібності (Similarity based Feature Selection), теорії інформації (Information Theoretical based), розрідженого навчання (Sparse Learning based), гібридні методи, глибокого навчання та на основі реконструкції, вбудовані підходи (Embedded approaches) та методи фільтрування та обгорток (Wrapper approaches) до вибору ознак, методи зважування ознак (Feature-weighting), алгоритм Winnow, алгоритми

для ознак групової структури (Feature Selection Algorithms with Group Structure Features), деревовидної структури (Feature Selection Algorithms with Tree Structure Features), та графової структури (Feature Selection Algorithms with Graph Structure Features), для неоднорідних даних (Feature selection with heterogeneous data): алгоритми для пов'язаних даних (Feature Selection Algorithms with Linked Data), ті що на основі множини джерел (Multi-Source Feature Selection) та множини поглядів (Multi-View Feature Selection), для потокових даних (Feature selection with streaming data): алгоритми для потоків ознак (Feature Selection Algorithms with Feature Streams) і для потоків даних (Feature Selection Algorithms with Data Streams), проблеми нерелевантних ознак та нерелевантних прикладів, засоби та показники оцінювання якості методів та алгоритмів вибору ознак;

теорію традиційного методу PCA, зростаючого та розрідженого методів головних компонент (Incremental PCA, SparsePCA and MiniBatchSparsePCA), ядерного методу головних компонент (kernel PCA), імовірнісного методу PCA, математичні основи факторного аналізу, словникового навчання (Dictionary Learning), методу невід'ємної матричної факторизації (Non-negative matrix factorization), методу аналізу незалежних компонент (Independent component analysis), теорію методу лінійного та квадратичного дискримінантного аналізу (Linear and Quadratic Discriminant Analysis), скритого розподілу Дирихле (Latent Dirichlet Allocation);

метод максимальної правдоподібності для оцінювання параметрів моделей, властивості оцінок максимальної правдоподібності, байесівський підхід до оцінювання параметрів моделей, теорема Байеса, максимальна апостеріорна гіпотеза, байесівський підхід до класифікації, оптимальний байесівський класифікатор, оцінювання апріорних імовірностей та функцій правдоподібності за вибіркою, наївний байесівський класифікатор;

загальних методів до регуляризації моделі, пошуку компромісу між систематичною помилкою і дисперсією моделі, засобів діагностування проблем зі зміщенням і дисперсією моделі, оцінювання якості моделей класифікації, вибору гіперпараметрів моделі;

методів та алгоритмів ієрархічної кластеризації, k-середніх, нечітких k-середніх та g-середніх, графових алгоритмів, FOREL та Expectation-Maximization з послідовним додаванням компонент, щільнісні алгоритми Mean Shift, DBSCAN, OPTICS та їх модифікації, методів конкурентного навчання, самоорганізуючих карт Кохонена, спектральної кластеризації, Affinity propagation та Birch, проблем оцінювання якості результатів кластеризації, різних підходів, методів та показників для оцінки якості моделей кластеризації;

структурних ймовірнісних моделей у глибокому навчанні, проблем безструктурного моделювання, графових структурних моделей: орієнтованих, неорієнтованих моделей, факторні графи, енергетичні моделі, методи вибірки і Монте-Карло, вибірки за значимістю, вибірки за Гіббсом;

методів наближеного виводу, MAP-виводу і розрідженого кодування, варіаційного виводу і навчання: на основі дискретних і неперервних латентних змінних, взаємодії між навчанням і виводом, методів навченого наближеного виводу;

загальних підходів до асоціативного аналізу (Association Mining) та аналізу шаблонів послідовностей (Sequential Pattern Mining), алгоритмів асоціативного аналізу: Apriori, Eclat, FP-growth та їх модифікацій, показники оцінювання якості асоціативних правил, алгоритм Apriori-ALL, алгоритми BFS (Breadth First Search) та DFS (Depth First Search) для аналізу шаблонів послідовностей, алгоритми GSP, SPADE, FreeSpan, PrefixSpan і SPAM, алгоритми в замкнутій формі для аналізу послідовностей: CloSpan і BIDE, інкрементні алгоритми ISM, IncSP, ISE, IncSpan та IncSpan+, MILE, критерії якості алгоритмів пошуку шаблонів послідовностей;

моделей понижуючого та регуляризованого автокодувальників, розрідженого, шумопригнічуючого, стохастичного кодувальника-декодувальника, марковської мережі, асоційованої з довільним шумоподавляючим автокодувальником, репрезентативної здатності автокодувальника, вибору розміру шару і глибини, методів вибірки з автокодувальників, методів навчання автокодувальників, теорії побудови варіаційного автокодувальника;

орієнтованих породжуючих моделей, теорії генеративних змагальних мереж (GAN), згорткових мереж DCGAN та інших модифікацій GAN, сигмоїдних мереж довіри, авторегресивних мереж;

різних видів рекомендаційних систем, проблем їх розробки, методів навчання ранжуванню, колаборативної фільтрації, нейронної колаборативної фільтрації, контентної фільтрації, алгоритм SVD, гібридних алгоритмів, теорії рекурентних нейронних мереж, алгоритму зворотного розповсюдження в часі (BackPropagation Throught Time) для навчання рекурентних мереж, моделей LSTM і GRU та їх модифікації, моделі рекомендаційних систем на основі рекурентних нейронних мереж типу кодувальник-декодувальник, методи і показники оцінювання якості рекомендацій;

технологій побудови, навчання та оцінювання якості моделей класифікації, кластеризації, асоціативних правил, аналізу шаблонів послідовностей, глибоких нейронних мереж прямого розповсюдження сигналу, моделей кодувальник-декодувальник, згорткових нейронних мереж, генеративних змагальних мереж, рекомендаційних систем в python з використанням бібліотек TensorFlow, Keras, scikit-learn, pandas, matplotlib, mlxtend.

уміння:

застосовувати описані вище сучасні моделі, методи і алгоритми інтелектуального аналізу великих сховищ даних і машинного навчання;

виконувати попередню обробку даних і побудову навчальних наборів, оцінювати важливість ознак і обирати значущі ознаки (feature selection);

зменшувати розмірність наборів даних (feature extraction) алгоритмами PCA, Incremental PCA, SparsePCA, MiniBatchSparsePCA, Kernel PCA, словникового навчання (Dictionary Learning), методу невід'ємної матричної факторизації (Non-negative matrix factorization), аналізу незалежних компонент (Independent component analysis), лінійного та квадратичного дискримінантного аналізу (Linear and Quadratic Discriminant Analysis), скритого розподілу Дирихле (Latent Dirichlet Allocation;

зменшувати розмірність наборів даних різними методами кластеризації;

будувати, навчати та оцінювати якість моделей кластеризації та класифікації;

шукати компроміс між систематичною помилкою і дисперсією моделі навчання з вчителем, діагностувати проблеми зі зміщенням і дисперсією моделей цього класу, обирати гіперпараметри моделей;

прогнозувати споживчий попит на основі наборів даних транзакцій алгоритмами асоціативного аналізу (association mining) та аналізу шаблонів послідовностей (sequential pattern mining), оцінювати якість побудованих асоціативних правил та послідовностей і прогнозів на їх основі;

виконувати прогнозування на основі глибоких нейромережевих моделей прямого розповсюдження сигналу та згорткових нейронних мереж, обирати функції активації, алгоритм оптимізації, параметри моделі, оцінювати якість навченої моделі, зберігати і повторно завантажувати навчені моделі;

здійснювати попереднє навчання класифікаторів з використанням глибоких автокодувальників; виконувати реконструкцію і породження зображень на основі глибоких моделей автокодувальників;

породжувати нові зображення моделями і методами генеративних змагальних мереж GAN, DCGAN та іншими модифікаціями мереж GAN;

будувати моделі рекомендаційних систем різними методами.

досвід:

теоретичний та практичний досвід аналізу і обробки даних у різних форматах з метою підтримки прийняття рішень, побудови прогнозів, породження нових даних, використання програмного забезпечення Python для інтелектуального аналізу даних та машинного навчання в практичній роботі.

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

При вивченні дисципліни використовуються знання дисциплін «Теорія ймовірностей», «Математична статистика», «Математичний аналіз», «Лінійна алгебра», «Методи оптимізації», «Чисельні методи», «Об'єктно-орієнтовне програмування», «Дискретна математика», «Математична логіка», «Інтелектуальний аналіз даних», «Інтелектуальні системи підтримки прийняття рішень», знають синтаксис мови програмування Python.

Знання, набуті при вивченні цієї дисципліни, використовуються в дипломному проектуванні, у практичній самостійній роботі випускника в галузі інтелектуального аналізу даних під час аналізу великих і надвеликих баз даних та масивів тексту, при побудові прогнозів на основі статистичних даних та оцінок експертів, при розробці корпоративних інформаційно-аналітичних систем в державних і приватних управлінських структурах.

3. Зміст навчальної дисципліни

Вступ. *Задачі ІАВСД*

Попередня обробка даних, вибір значущих ознак, зниження розмірності даних, сегментація, кластеризація, класифікація, породження нових даних, структурний вивід, наближений вивід, синтез і вибірка, асоціативний аналіз, навчання ранжуванню, побудова рекомендаційних систем. Огляд методів ІАД та машинного навчання.

Розділ 1. Попередня обробка даних і побудова навчальних наборів

Тема 1.1. Обробка категоріальних даних.

Тема 1.2. Розв'язання проблеми з відсутніми даними: їх ідентифікація в таблицях, підходи до розрахунку даних, що відсутні.

Тема 1.3. Оцінювання важливості ознак і вибір значущих ознак (Feature Selection).

Розділ 2. Зниження розмірності набору даних, отримання суттєвих ознак (Feature Extraction)

Тема 2.1. Метод головних компонент (PCA) для зниження розмірності без вчителя. Ядерний метод головних компонент (Kernel PCA).

Тема 2.2. Імовірнісний метод PCA. Факторний аналіз.

Тема 2.3. Словникове навчання (Dictionary Learning).

Тема 2.4. Метод невід'ємної матричної факторизації (Non-negative matrix factorization).

Тема 2.5. Аналіз незалежних компонент (Independent component analysis).

Тема 2.6. Лінійний та квадратичний дискримінантний аналіз (Linear and Quadratic Discriminant Analysis).

Тема 2.7. Скритий розподіл Дирихле (Latent Dirichlet Allocation).

Розділ 3. Оцінка якості алгоритмів навчання з вчителем

Тема 3.1. Перенавчання (overfitting) моделі. Регуляризація моделі. Перехресна перевірка (cross-validation, CV) та її модифікації StratifiedKFold, ShuffleSplit, StratifiedShuffleSplit, LeaveOneOut. Ітератори перехресної перевірки. Діагностування проблем зі зміщенням і дисперсією моделі. Вибір гіперпараметрів моделі методами решітчастого Grid Search CV та рандомізованого пошуку Random Search CV.

Тема 3.2. Метод максимальної правдоподібності для оцінювання параметрів моделей. Властивості оцінок максимальної правдоподібності. Байесівський підхід до оцінювання параметрів моделей. Теорема Байеса. Максимальна апостеріорна гіпотеза. Байесівський підхід до класифікації.

Тема 3.3. Оцінювання якості моделей класифікації. Матриця неточностей (confusion matrix). Показники accuracy, precision, recall, specificity, F1-score для вибору моделі. Криві ROC-curve та PR-curve. Розрахунок показників якості в задачі багатокласової класифікації. Проблема незбалансованих класів та шляхи її вирішення.

Розділ 4. Кластеризація даних різної форми алгоритмами навчання без вчителя

Тема 4.1. Ієрархічна кластеризація.

Тема 4.2. Алгоритми k-середніх, нечітких k-середніх та g-середніх.

Тема 4.3. Графові алгоритми кластеризації. Алгоритм FOREL. Алгоритм Expectation-Maximization (EM) з послідовним додаванням компонент.

Тема 4.4. Щільнісні алгоритми кластеризації Mean Shift, DBSCAN, OPTICS та їх модифікації. Реалізація алгоритмів та вибір гіперпараметрів. Аналіз результатів кластеризації наборів даних різної форми.

Тема 4.5. Самоорганізуючі карти Кохонена. Мережі Кохонена. Конкурентне навчання. Інтерпретація карт Кохонена.

Тема 4.6. Методи спектральної кластеризації, Affinity propagation та Birch.

Тема 4.7. Оцінка якості та аналіз результатів кластеризації.

Розділ 5. Структурні ймовірнісні моделі у глибокому навчанні

Тема 5.1. Проблема безструктурного моделювання. Застосування графів для описання структури моделі.

Тема 5.2. Вибірка і методи Монте-Карло. Вибірка за значимістю. Методи Монте-Карло за схемою марковської мережі. Вибірка за Гіббсом.

Тема 5.3. Наближений вивід. Вивід як оптимізація. MAP-вивід і розріджене кодування. Варіаційний вивід і навчання: дискретні і неперервні латентні змінні. Взаємодія між навчанням і виводом. Навчений наближений вивід.

Розділ 6. Асоціативний аналіз (Association Mining) та аналіз шаблонів послідовностей (Sequential Pattern Mining)

Тема 6.1. Задача пошуку асоціативних правил та аналізу ринкових кошиків. Загальний підхід до її розв'язання. Метрики оцінювання якості асоціативних правил. Алгоритми асоціативного аналізу: Apriori, Eclat та їх реалізація на python.

Тема 6.2. Алгоритми FP-growth. Реалізація цих алгоритмів на python.

Тема 6.3. Алгоритми пошуку шаблонів послідовностей та їх реалізація на python.

Розділ 7. Моделі типу кодувальник-декодувальник. Автокодувальники

Тема 7.1. Понижуючі, регуляризовані, стохастичні, варіаційні автокодувальники. Репрезентативна здатність, розмір шару і глибина.

Тема 7.2. Вибірка з автокодувальників. Марковська мережа, асоційована з довільним шумоподавляючим автокодувальником.

Тема 7.3. Застосування автокодувальників. Побудова і навчання автокодувальників в TensorFlow.

Розділ 8. Орієнтовані породжуючі моделі

Тема 8.1. Основи теорії генеративних змагальних мереж (GAN). Побудова мереж GAN в TensorFlow. Застосування GAN.

Тема 8.2. Згорткові мережі GAN – мережі DCGAN. Мережі GAN Вассерштейна. Реалізація в TensorFlow. Застосування.

Тема 8.3. Генеративна змагальна мережа найменших квадратів (LSGAN).

Тема 8.4. Інші модифікації GAN.

Тема 8.5. Сигмоїдні мережі довіри. Авторегресивні мережі.

Розділ 9. Рекомендаційні системи

Тема 9.1. Види рекомендаційних систем. Навчання ранжуванню. Колаборативна фільтрація. Алгоритм SVD. Проблеми розробки рекомендаційних систем.

Тема 9.2. Метод контентної фільтрації.

Тема 9.3. Метрики оцінювання якості рекомендацій.

Тема 9.4. Гібридні алгоритми побудови прогнозу в рекомендаційних системах. Нейронна колаборативна фільтрація.

Тема 9.5. Основи рекурентних нейронних мереж. Модель в просторі станів. Задачі обробки послідовностей. Алгоритми зворотного розповсюдження в часі (BackPropagation Throught Time) для навчання рекурентних мереж. Проблеми навчання рекурентних нейронних мереж.

Тема 9.6. Модель довгої короткотермінової пам'яті (Long Short-Term Memory, LSTM). Модель GRU. Модифікації LSTM.

Тема 9.7. Моделі рекомендаційних систем на основі рекурентних нейронних мереж типу кодувальник-декодувальник. Реалізація в TensorFlow.

Напрямки розвитку та перспективи подальших досліджень в області ІА великих сховищ даних та машинного навчання. Невирішені проблеми.

4. Навчальні матеріали та ресурси

Базова

1. *Н.І. Недашківська. Конспект лекцій у формі слайдів з кредитного модуля «Інтелектуальний аналіз великих сховищ даних», 2024, https://classroom.google.com/c/NjYxOTMwNTA2OTI2?cjc=mqj7y7u*
2. *Н.І. Недашківська. Методичні вказівки до виконання практичних робіт з кредитного модуля «Інтелектуальний аналіз великих сховищ даних», 2024, https://classroom.google.com/c/NjYxOTMwNTA2OTI2?cjc=mqj7y7u*
3. *Н.І. Недашківська. Інтелектуальний аналіз даних : Практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 124 «Системний аналіз», освітніх програм «Системний аналіз і управління», «Системний аналіз фінансового ринку»/ Н. І. Недашківська; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 6 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2021. – 105 с. https://ela.kpi.ua/handle/123456789/53763*
4. *Н.І. Недашківська. Методи і моделі інтелектуального аналізу даних: Практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 122 «Комп'ютерні науки», освітньої програми «Системи і методи штучного інтелекту» / Н. І. Недашківська; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 1,8 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2019. – 70 с. https://ela.kpi.ua/handle/123456789/53764*

Додаткова література

5. *Scikit-Learn Documentation. Режим доступу: https://scikit-learn.org/, 2024.*
6. *TensorFlow Documentation. Режим доступу: https://www.tensorflow.org . 2024.*
7. *Keras Documentation. Режим доступу: https://keras.io. 2024.*
8. *Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning. The MIT Press Cambridge, Massachusetts London, England, 2017. https://www.deeplearningbook.org/*
9. *Sebastian Raschka, Vahid Mirjalili. Python Machine Learning. Third Edition. Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2. Packt Publishing, 2019. https://github.com/rasbt/python-machine-learning-book-3rd-edition*
10. *Jake VanderPlas. Python Data Science Handbook. Essential Tools for Working with Data. O'Reilly Media Inc., 2017. 576 p. (за запитом викладачу)*
11. *Aurelien Geron. Hands-On Machine Learning with Scikit-Learn and TensorFlow. O'Reilly Media Inc., Sebastopol, CA, 2017. (за запитом викладачу)*
12. *Rodolfo Bonnin. Building Machine Learning Projects with TensorFlow. Packt Publishing Ltd., Birmingham, Uk, 2016. (за запитом викладачу)*
13. *Ramsundar B., Zadeh R.B.. TensorFlow for Deep Learning. O'Reilly Media Inc., Sebastopol, CA, 2018. (за запитом викладачу)*
14. *Wes McKinney. Python for Data Analysis. O'Reilly Media Inc., 2013. 482 p. (за запитом викладачу)*
15. *Henrik Brink Joseph W. Richards Mark Fetherolf. Real-World Machine Learning. Manning Publications. 2016. 336 p. (за запитом викладачу)*
16. *Andreas C. Mueller and Sarah Guido. An Introduction to Machine Learning with Python. O'Reilly Media Inc., 2017. 392 p. (за запитом викладачу)*
17. *Davy Cielen, Arno Meysman, Mohamed Ali. Introducing Data Science: Big Data, Machine Learning, and more, using Python tools. Manning Publications, 2016. 320 p. (за запитом викладачу)*

Використовується сучасне комп'ютерне та мультимедійне обладнання, платформа дистанційного навчання «Сікорський».

Для виконання практичних робіт використовується open-source програмне забезпечення Python (https://www.python.org/), Scikit-Learn 1.2.1 – open source, commercially usable – BSD license (https://scikit-learn.org/), TensorFlow v.2.11.0 – Apache-2.0 license (https://www.tensorflow.org), Keras – Apache-2.0 license (https://keras.io)

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Лекційні заняття

Лекція 1 (U-1). Загальні відомості про ІАД. Досвід в задачах ІАД. Задачі ІАД

Загальні відомості про ІАД. Досвід в задачах ІАД: навчання з вчителем – supervised learning, без вчителя – unsupervised learning, з частковим залученням вчителя – semi-supervised learning. Задачі ІАД: попередня обробка даних, вибір значущих ознак, зниження розмірності даних, сегментація, кластеризація, класифікація, породження нових даних, структурний вивід, наближений вивід, синтез і вибірка, асоціативний аналіз, навчання ранжуванню, побудова рекомендаційних систем. Огляд методів ІАД та машинного навчання.

Лекція 2 (U-2). Розв'язання проблеми з відсутніми даними. Обробка категоріальних даних

Обробка категоріальних даних: номінальні і порядкові ознаки, відображення порядкових ознак за допомогою pandas, кодування міток класів, one-hot кодування номінальних ознак, кодування порядкових ознак.

Розв'язання проблеми з відсутніми даними: ідентифікація відсутніх значень у табличних даних, вилучення навчальних прикладів чи ознак з відсутніми значеннями, заповнення відсутніх значень шляхом інтерполяції: одновимірний алгоритм (univariate feature imputation), багатовимірний алгоритм (multivariate feature imputation). Заповнення пропущених значень перед побудовою оцінювача. Дослідження різних варіантів багатовимірного алгоритму IterativeImputer.

Лекція 3 (U-3). Методи оцінювання важливості ознак (Feature Selection)

Традиційна категоризація алгоритмів вибору ознак: з точки зору наявності вчителя (Supervision Perspective), з точки зору різних стратегій вибору (Selection Strategy Perspective). Алгоритми вибору ознак з точки зору даних (from a Data Perspective).
Явище перенавчання моделі. Регуляризація L1 і L2 - підхід до зменшення складності моделі шляхом штрафування великих індивідуальних ваг. Геометрична інтерпретація L1 і L2 регуляризації. Розріджені рішення з L1-регуляризацією. Вибір ознак засобами L1-регуляризації. Алгоритми послідовного вибору ознак (sequential feature selection, SFS) та зниження розмірності набору даних. Алгоритм послідовного зворотного вибору (sequential backward selection, SBS). Алгоритми рекурсивного виключення ознак (recursive feature elimination, RFE) та RFECV. Оцінювання важливості ознак на основі моделі класифікації / регресії за допомогою ridge, lasso, LinearSVC, логістичної регресії, випадкових лісів.

Лекція 4 (U-4). Методи оцінювання важливості і вибору ознак (частина 2)

Традиційні методи вибору ознак для традиційних даних (Traditional Feature Selection for Conventional Data): на основі подібності (Similarity based Feature Selection Methods),
теорії інформації (Information Theoretical based Feature Selection Methods),
розрідженого навчання (Sparse Learning based Feature Selection Methods).
Статистичні методи вибору ознак (Statistical based Feature Selection Methods).
Гібридні методи, методи на основі глибокого навчання та на основі реконструкції.

Лекція 5 (U-5). Методи оцінювання важливості і вибору ознак (частина 3)

Вбудовані підходи (Embedded approaches) до вибору ознак. Методи фільтрування та обгорток (Wrapper approaches) для вибору ознак.
Методи зважування ознак (Feature-weighting). Алгоритм Winnow.
Проблема нерелевантних ознак. Визначення поняття «релевантність». Вибір ознак як евристичний пошук. Проблема нерелевантних прикладів (irrelevant examples). Вибір розмічених даних (labeled data). Вибір немаркованих даних (unlabeled data).
Оцінювання методів та алгоритмів вибору ознак. Показники оцінювання якості. Відкриті проблеми: масштабованість, стабільність, вибір моделі.

Лекція 6 (U-6). Метод головних компонент (PCA) для зниження розмірності даних без вчителя

Теорія традиційного методу PCA, зростаючого та розрідженого методів головних компонент (Incremental PCA, SparsePCA and MiniBatchSparsePCA). Ядерний метод головних компонент (kernel PCA).
Метод головних компонент в Scikit-learn Python. Реалізація ядерного методу головних компонент на Python. Ядерний метод головних компонент в scikit-learn python. Приклади.

Лекція 7 (U-7). Імовірнісний метод PCA

Математичні основи методу PCA з максимальною правдоподібністю. Математичні основи факторного аналізу.

Імовірнісний метод PCA в scikit-learn python. Алгоритми Factor Analysis та Independent component analysis в scikit-learn python. Приклади.

Лекція 8 (U-8). Словникове навчання (Dictionary Learning). Метод невід'ємної матричної факторизації (Non-negative matrix factorization)

Математичні основи словникового навчання. Математичні основи методу невід'ємної матричної факторизації.

Алгоритми Dictionary Learning та MiniBatchDictionaryLearning в scikit-learn python. Алгоритми Non-negative matrix factorization в scikit-learn python. Приклади.

Лекція 9 (U-9). Оцінка моделей і налаштування гіперпараметрів

Перехресна перевірка (cross-validation, CV), KFold. Перенавчання (overfitting) моделі. Регуляризація моделі. Компроміс між систематичною помилкою і дисперсією моделі. Діагностування проблем зі зміщенням і дисперсією моделі. Вибір гіперпараметрів моделі методами решітчастого Grid Search CV та рандомізованого пошуку Random Search CV.

Метод максимальної правдоподібності для оцінювання параметрів моделей. Властивості оцінок максимальної правдоподібності. Байесівський підхід до оцінювання параметрів моделей. Теорема Байеса. Максимальна апостеріорна гіпотеза.

Байесівський підхід до класифікації. Оптимальний байесівський класифікатор. Оцінювання апріорних імовірностей та функцій правдоподібності за вибіркою. Наївний байесівський класифікатор. Приклади розв'язання задач.

Лекція 10 (U-10). Оцінювання якості моделей класифікації

Матриця неточностей (confusion matrix). Показники accuracy, precision, recall, specificity, F1-score для вибору моделі. Криві ROC-curve та PR-curve. Розрахунок показників якості в задачі багатокласової класифікації. Проблема незбалансованих класів та шляхи її вирішення.

Перехресна перевірка KFold та її модифікації StratifiedKFold, ShuffleSplit, StratifiedShuffleSplit, LeaveOneOut. Iтератори перехресної перевірки.

Лекція 11 (U-11). Ієрархічна кластеризація

Вступ до методів кластеризації. Функції відстані. Ієрархічна кластеризація: агломеративний алгоритм найближчого сусіда AgglomerativeClustering. Ієрархічна кластеризація: дівізимний алгоритм DIANA. Алгоритм AgglomerativeClustering scikit-learn python. Методи розрахунку відстані між кластерами. Приклад кластеризації наборів даних різної форми алгоритмом AgglomerativeClustering scikit-learn python. Приклад побудови дендрограми. Різні методи розрахунку відстані між кластерами: порівняльний аналіз результатів.

Лекція 12 (U-12). Алгоритми k-середніх, нечітких k-середніх та g-середніх

Базовий алгоритм KMeans. Алгоритми нечітких k-середніх, fuzzy KMeans, g-середніх, GMeans. Алгоритм MiniBatch KMeans. Реалізація в scikit-learn python: алгоритми KMeans, MiniBatchKMeans.

Емпірична оцінка впливу ініціалізації в методі k-середніх. Порівняння алгоритмів KMeans та MiniBatchKMeans на наборах даних make_blobs. Кластеризація текстових документів методом k-середніх. Порівняння результатів алгоритмів Bisecting K-Means та Regular K-Means на наборі даних make_blobs.

Лекція 13 (U-13). Графові алгоритми кластеризації. Алгоритм FOREL. Алгоритм Expectation-Maximization (EM) з послідовним додаванням компонент

Методи кластеризації на основі теорії графів. Алгоритм знаходження зв'язних компонент. Поняття мінімального покриваючого дерева (МПД). Жадібний алгоритм побудови МПД. Алгоритми Прима, Крускала і Борувки побудови МПД. Приклади

Базовий FOREL, модифікації FOREL - 2, 3, 4 для опису даних складної форми. Вибір кількості кластерів. Приклади

Задача розділу суміші. Алгоритм EM з фіксованою кількістю компонент. Недоліки базового алгоритму EM. Модифікації алгоритму EM: узагальнений, стохастичний. Модифікований алгоритм EM з послідовним додаванням компонент для розв'язання задачі кластеризації.

Лекція 14 (U-14). Щільнісні алгоритми кластеризації

Математичні основи методів Mean Shift, DBSCAN, OPTICS. Модифікації методів. Опис алгоритмів, їх реалізація та вибір гіперпараметрів.

Алгоритми Mean Shift, DBSCAN, OPTICS в scikit-learn python. Приклади. Аналіз результатів кластеризації наборів даних різної форми.

Лекція 15 (U-15). Метод самоорганізуючих карт Кохонена

Задача оптимізації для розрахунку центрів кластерів. Використання методу стохастичного градієнтного спуску. Нейронна мережа Кохонена для кластеризації. Алгоритм розрахунку центрів кластерів. Конкурентне навчання. Жорстке і м'яке правило розрахунку центрів кластерів. Карта Кохонена – прямокутна або шестигранна сітка кластерів. Метод самоорганізуючих карт Кохонена. Алгоритм навчання карти Кохонена. Інтерпретація карт Кохонена.

Приклад сегментації абонентської бази білінгової системи телекомунікаційної компанії на основі статистики використаних послуг.

Лекція 16 (U-16). Методи спектральної кластеризації, Affinity propagation та Birch

Математичні основи методів Spectral Clustering, Affinity propagation та Birch. Опис алгоритмів, їх реалізація та вибір гіперпараметрів.

Алгоритми SpectralClustering, AffinityPropagation та Birch в scikit-learn python. Приклади. Аналіз результатів кластеризації наборів даних різної форми.

Лекція 17 (U-17). Оцінка якості та аналіз результатів кластеризації

Порівняння результатів на даних складної форми, стійкість до шумів, швидкодія алгоритмів. Рекомендовані етапи кластерного аналізу. Відносна та внутнішня валідація. Методи ресемплінгу.

Коефіцієнт силуету (silhouette). Вибір кількості кластерів за допомогою аналізу коефіцієнтів силуету. Оцінювання наявності кластерів у заданій вибірці за статистикою Хопкінса. Індекс Calinski-Harabasz та Davies-Bouldin.

Показники якості на основі додатково відомих розмічених даних:

- *однорідність (homogeneity), повнота (completeness), v-measure,*
- *коефіцієнт розбиття, індекс чіткості,*
- *індекс Ренда (Rand index), adjusted Rand index,*
- *показники на основі взаємної інформації (mutual information): normalized mutual information, adjusted mutual information,*
- *індекс Фоулкса – Меллова (Fowlkes-Mallows),*
- *матриця випадковості (Contingency Matrix),*
- *попарна матриця неточностей (Pair Confusion Matrix).*

Переваги і недоліки різних показників якості. Оцінка якості кластеризації в scikit-learn python.

Лекція 18 (U-18). Алгоритми Apriori та Eclat

Постановка задачі аналізу ринкових кошиків. Поняття асоціативного правила. Підтримка набору. Властивість антимонотонності. Показники корисності асоціативних правил.

Алгоритми Apriori, їх переваги, недоліки. Підходи до підвищення ефективності Apriori. Алгоритм Eclat. Приклади застосування.

Лекція 19 (U-19). Алгоритм FP-growth

Поняття префіксного дерева (FP-дерево, frequent pattern tree). Алгоритм побудови FP-дерева. Приклад.

Пошук частих наборів в FP-дереві: поняття умовного FP-дерева, алгоритм побудови умовного FP-дерева, алгоритм пошуку частих наборів в FP-дереві.

Переваги і недоліки алгоритму FP-росту. Порівняння алгоритмів Apriori та FP-росту. Приклади застосування. Модифікації алгоритму FP-росту.

Лекція 20 (U-20). Алгоритми пошуку шаблонів послідовностей (Sequential Pattern Mining)

Періодичні (Periodic Patterns), статистично значущі патерни та орієнтовні патерни (Approximate Patterns).

Аналіз послідовностей алгоритмом на основі Apriori-ALL (Apriori-like algorithm).

Алгоритми BFS (Breadth First Search) для аналізу послідовностей. Алгоритм GSP. Алгоритми DFS (Depth First Search) для аналізу послідовностей. Алгоритми SPADE, FreeSpan, PrefixSpan і SPAM.

Критерії якості алгоритмів пошуку шаблонів послідовностей.

Лекція 21 (U-21). Автокодувальники (частина 1)

Поняття автокодувальника та його складові. Навчання детермінованого автокодувальника. Понижуючий (undercomplete) автокодувальник для зниження розмірності даних. Регуляризований автокодувальник. Нагадування про регуляризацію на прикладі поліноміальної регресії. Регуляризована функція втрат. Обгрунтування регуляризованої функції втрат, використовуючи теорію Байеса. Розріджений (sparse) автокодувальник. Породжуюча модель.

Лекція 22 (U-22). Автокодувальники (частина 2)

Розріджений (sparse) автокодувальник. Моделювання розрідженості. Втрата внаслідок розрідженості - регуляризуючий член на основі розходження Кульбака - Лейблера. Шумопригнічуючий автокодувальник. Два способи введення шуму в модель.

Глибокі автокодувальники. Зв'язування ваг. Два способи навчання глибокого автокодувальника. Породження даних на основі детермінованого автокодувальника.

Попереднє навчання без учителя з використанням глибоких автокодувальників.

Теорія побудови варіаційного автокодувальника.

Лекція 23 (U-23). Автокодувальники (частина 3)

Реалізація моделей автокодувальників з нуля в низькорівневому API Tensorflow 1.x Python. Зменшення розмірності даних, використовуючи понижуючий (undercomplete) автокодувальник.

Побудова глибокого (stacked, deep) автокодувальника двома способами. Використання глибокого автокодувальника для реконструкції зображень MNIST. Реалізація зв'язування ваг. Відображення ознак, отриманих автокодувальником.

Використання шарів попередньо навченого глибокого автокодувальника для побудови класифікатора набору зображень MNIST. Порівняння результатів з тими, що були отримані традиційним способом навчання.

Побудова шумопригнічуючого (denoising) автокодувальника. Два способи введення шуму в модель. Використання шумопригнічуючого автокодувальника для реконструкції зображень MNIST.

Побудова розрідженого (sparse) автокодувальника. Використання розрідженого автокодувальника для реконструкції зображень MNIST.

Побудова варіаційного автокодувальника. Використання варіаційного автокодувальника для генерування зображень, які виглядають як рукописні цифри MNIST.

Лекція 24 (U-24). Основи генеративних змагальних мереж (GAN)

Поняття породжуючої моделі. Поняття генеративної змагальної мережі. Генератор і дискримінатор - складові генеративної змагальної мережі (GAN). Функції втрат генератора і дискримінатора.

Налаштування середовища Google Colab для навчання моделей GAN. Реалізація простої моделі GAN з нуля. Попередня обробка набору зображень MNIST перед подачею на вхід моделі GAN. Оцінювання якості моделі GAN. Побудова мереж GAN в TensorFlow. Застосування GAN.

Лекція 25 (U-25). Згорткові GAN (DCGAN) та модель GAN Вассерштейна

Генерування зображень кращої якості використовуючи згорткові GAN та модель GAN Вассерштейна. Поняття транспонованої згортки (Transposed convolution). Шар номалізації за батчами (Batch normalization).

Побудова моделі DCGAN з нуля: модель генератора, модель дискримінатора.

Міри відмінності між двома розподілами: розходження Кульбака-Лейблера, розходження Йенсена-Шенона, відстань Earth mover's.

Налаштування GPU на Google Colab. Модель WGAN-GP, її побудова і навчання. Використання Gradient penalty (GP) та відстані Earth mover's. Налаштування набору даних MNIST. Обчислення функцій втрат генератора і дискримінатора. Обчислення Gradient penalty.

Напрямки розвитку та перспективи подальших досліджень в області ІА великих сховищ даних та машинного навчання. Невирішені проблеми.

Практичні роботи

Метою практичних робіт є закріплення теоретичних положень навчальної дисципліни, отримання практичних навичок створення, навчання і оцінювання якості моделей інтелектуального аналізу великих сховищ даних.

№ з/п	Назва роботи	Кількість ауд. годин
1	*Заповнення відсутніх значень в наборах даних алгоритмами інтерполяції.*	2
2	*Оцінювання важливості ознак і вибір значущих ознак (feature selection) засобами scikit-learn python.*	4
3	*Дослідження методів feature selection для структурованих ознак, неоднорідних і потокових даних.*	4
4	*Зниження розмірності набору даних, отримання суттєвих ознак (feature extraction) методом головних компонент (PCA).*	4
5	*Зниження розмірності набору даних методами ICA, FA, NNMF, Dictionary Learning, LDA.*	4
6	*Кластеризація даних різної форми алгоритмами навчання без вчителя. Оцінка якості результатів кластеризації.*	4
7	*Прогнозування на основі наборів даних транзакцій алгоритмами Apriori, Eclat, FP-growth. Оцінка якості побудованих асоціативних правил та прогнозу.*	2
8	*Аналіз шаблонів послідовностей алгоритмами BFS (Breadth First Search-based) та DFS (Depth First Search-based). Оцінка якості результатів.*	2
9	*Аналіз шаблонів послідовностей інкрементними алгоритмами та алгоритмами в замкнутій формі. Оцінка якості результатів.*	2
10	*Реконструкція і породження зображень на основі глибоких моделей автокодувальників.*	2
11	*Породження зображень змагальними мережами GAN, DCGAN та GAN Вассерштейна.*	2
12	*Породження зображень змагальними мережами GAN, DCGAN та GAN Вассерштейна.*	4

Для виконання практичних робіт використовується open-source програмне забезпечення Python (https://www.python.org/), Scikit-Learn 1.2.1 – open source, commercially usable – BSD license (https://scikit-learn.org/), TensorFlow v.2.11.0 – Apache-2.0 license (https://www.tensorflow.org), Keras – Apache-2.0 license (https://keras.io)

6. Самостійна робота студента

Самостійна робота студента включає підготовку до практичних/ лабораторних робіт, підготовку до модульної контрольної роботи, в тому числі опрацювання наступних матеріалів по темам силабусу.

Тема 1.1. *Кодування порядкових ознак.*

Тема 1.2. *Багатовимірні алгоритми заповнення відсутніх значень у табличних даних (multivariate feature imputation).*

Параметри, атрибути і методи класів: sklearn.impute.SimpleImputer, sklearn.impute.IterativeImputer, sklearn.impute.KNNImputer.

Тема 1.3. *Вибір ознак для структурованих ознак: алгоритми для ознак групової структури (Feature Selection Algorithms with Group Structure Features), алгоритми для ознак деревовидної структури (Feature Selection Algorithms with Tree Structure Features), алгоритми для ознак графової структури (Feature Selection Algorithms with Graph Structure Features).*

Вибір ознак для неоднорідних даних (Feature selection with heterogeneous data): алгоритми для пов'язаних даних (Feature Selection Algorithms with Linked Data), алгоритми на основі множини джерел (Multi-Source Feature Selection) та на основі множини поглядів (Multi-View Feature Selection).

Вибір ознак для потокових даних (Feature selection with streaming data): алгоритми для потоків ознак (Feature Selection Algorithms with Feature Streams) і для потоків даних (Feature Selection Algorithms with Data Streams).

Пакет mlxtend python - реалізація кількох варіантів послідовного вибору ознак: алгоритми SequentialFeatureSelector.

Алгоритми модуля Feature selection в scikit-learn.

Тема 2.1. *Застосування методу головних компонент (PCA) до набору даних з ірисами Фішера.*

Клас sklearn.decomposition.IncrementalPCA, який реалізує зростаючий метод головних компонент (Incremental PCA). Метод IPCA призначений для великих наборів даних, які не вміщуються в основну пам'ять, що вимагає поступових підходів.

Порівняння результатів Incremental PCA та традиційного PCA на простому наборі з ірисами Фішера.

Клас sklearn.decomposition.SparsePCA, який реалізує розріджений метод головних компонент (SparsePCA and MiniBatchSparsePCA).

Застосування різних методів зниження розмірності даних без вчителя з модуля sklearn.decomposition до набору зображень The Olivetti faces dataset.

Використання KernelPCA для усунення шумів у зображеннях на прикладі набору цифр USPS. Навчання функції апроксимації для наступної реконструкції початкового зображення. Порівняння результатів з точною реконструкцією за допомогою PCA.

Тема 2.2. *Математичні основи факторного аналізу.*

Застосування алгоритмів імовірнісного PCA і факторного аналізу для оцінювання джерел з зашумлених вимірів.

Зниження розмірності різними методами набору даних Faces dataset decompositions.

Тема 2.4. *Клас sklearn.decomposition.MiniBatchNMF.*

Застосування різних методів зниження розмірності даних без вчителя з модуля sklearn.decomposition до набору зображень The Olivetti faces dataset.

Тема 2.5. *Теорія методу аналізу незалежних компонент (Independent component analysis). Клас sklearn.decomposition.FastICA, який реалізує швидкий алгоритм аналізу незалежних компонент.*

Порівняння результатів методом аналізу незалежних компонент (ICA) та PCA на штучно сгенерованих даних.

Тема 2.6. *Теорія методу лінійного та квадратичного дискримінантного аналізу (Linear and Quadratic Discriminant Analysis).*

Порівняння результатів лінійного дискримінантного аналізу (Linear Discriminant Analysis, LDA) та традиційного PCA на наборі з ірисами Фішера.

Тема 2.7. *Теорія методу скритого розподілу Дирихле (Latent Dirichlet Allocation).*

Застосування різних методів зниження розмірності даних без вчителя з модуля sklearn.decomposition до набору зображень The Olivetti faces dataset.

Тема 3.1. *Параметри, атрибути і методи класів RepeatedKFold, StratifiedKFold, ShuffleSplit, StratifiedShuffleSplit, LeaveOneOut.*

Порівняльний аналіз різних методів перехресної перевірки.

Тема 3.2. *Реалізація алгоритму Naive Bayes в scikit-learn python.*

Класи sklearn.naive_bayes.GaussianNB та sklearn.naive_bayes.MultinomialNB.

Приклади задач класифікації алгоритмом Naive Bayes: імовірнісне калібрування класифікаторів, криві імовірнісного калібрування.

Порівняльний аналіз алгоритмів байесівської класифікації.

Тема 4.1. *Приклади кластеризації наборів даних різної форми, використовуючи алгоритм AgglomerativeClustering, вибір значень гіперпараметрів.*

Тема 4.2. *Приклади кластеризації наборів даних різної форми, використовуючи алгоритми KMeans та MiniBatchKMeans, вибір значень гіперпараметрів.*

Тема 4.3. *Порівняльний аналіз результатів отриманих графовими алгоритмами кластеризації, FOREL та EM з послідовним додаванням компонент.*

Тема 4.4. *Порівняльний аналіз результатів отриманих алгоритмами Mean Shift, DBSCAN, OPTICS та їх модифікаціями.*

Тема 4.6. *Порівняльний аналіз результатів отриманих алгоритмами SpectralClustering, AffinityPropagation та Birch.*

Тема 4.7. *Експерименти з дослідження та коригування випадковості при оцінці продуктивності кластеризації. Перший експеримент: істинні мітки класів фіксовано, кількість кластерів збільшується. Другий експеримент: зміна кількості класів і кластерів.*

Тема 5.1. *Проблема безструктурного моделювання. Застосування графів для описання структури моделі: орієнтовані моделі, неорієнтовані моделі, факторні графи, енергетичні моделі. Переваги структурного моделювання.*

Тема 5.2. *Вибірка і методи Монте-Карло. Вибірка за значимістю. Методи Монте-Карло за схемою марковської мережі. Вибірка за Гіббсом.*

Тема 5.3. *Наближений вивід. Вивід як оптимізація. MAP-вивід і розріджене кодування. Варіаційний вивід і навчання: дискретні і неперервні латентні змінні. Взаємодія між навчанням і виводом. Навчений наближений вивід.*

Тема 6.2. *Модифіковані алгоритми FP-growth та їх реалізація.*

Тема 6.3. *Алгоритми в замкнутій формі для аналізу послідовностей. Алгоритми CloSpan і BIDE.*

Інкрементні алгоритми аналізу послідовностей: ISM, IncSP, ISE, IncSpan та IncSpan+, MILE. Реалізація алгоритмів.

Порівняльний аналіз різних алгоритмів пошуку шаблонів послідовностей (Sequential Pattern Mining).

Тема 7.1. *Стохастичні автокодувальники. Репрезентативна здатність автокодувальника, розмір шару і глибина. Сжимаючі автокодувальники. Модель прогнозної розрідженої декомпозиції (predictive sparse decomposition).*

Тема 7.2. *Вибірка з автокодувальників. Марковська мережа, асоційована з довільним шумоподавляючим автокодувальником.*

Тема 7.3. *Застосування автокодувальників. Побудова і навчання автокодувальників в TensorFlow. Співставлення рейтингів на основі шумоподавляючого автокодувальника. Навчання багатовидів за допомогою автокодувальників.*

Тема 8.1. *Застосування GAN.*

Тема 8.3*. Модифікації GAN та їх застосування.*

Тема 8.4. *Сигмоїдні мережі довіри. Авторегресивні мережі.*

Тема 9.1. *Види рекомендаційних систем. Навчання ранжуванню. Колаборативна фільтрація. Алгоритм SVD. Проблеми розробки рекомендаційних систем.*

Тема 9.2. *Метод контентної фільтрації.*

Тема 9.3. *Метрики оцінювання якості рекомендацій в різних видах рекомендаційних систем.*

Тема 9.4. *Гібридні алгоритми побудови прогнозу в рекомендаційних системах та їх реалізація. Нейронна колаборативна фільтрація.*

Тема 9.5. *Основи рекурентних нейронних мереж. Модель в просторі станів. Задачі обробки послідовностей. Алгоритми зворотного розповсюдження в часі (BackPropagation Throught Time) для навчання рекурентних мереж. Проблеми навчання рекурентних нейронних мереж.*

Тема 9.6. *Модель довгої короткотермінової пам'яті (Long Short-Term Memory, LSTM). Модель GRU. Модифікації LSTM.*

Тема 9.7. *Моделі рекомендаційних систем на основі рекурентних нейронних мереж типу кодувальник-декодувальник та їх реалізація.*

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Пропущені контрольні заходи оцінювання*. Кожен студент має право відпрацювати пропущені з поважної причини (лікарняний, мобільність тощо) заняття за рахунок самостійної роботи. Детальніше за посиланням: https://kpi.ua/files/n3277.pdf.*
Процедура оскарження результатів контрольних заходів оцінювання*. Студент може підняти будь-яке питання, яке стосується процедури контрольних заходів та очікувати, що*

воно буде розглянуто згідно із наперед визначеними процедурами. Студенти мають право аргументовано оскаржити результати контрольних заходів, пояснивши з яким критерієм не погоджуються відповідно до оціночного.

***Календарний контроль** проводиться з метою підвищення якості навчання студентів та моніторингу виконання студентом вимог силабусу.*

***Академічна доброчесність.** Політика та принципи академічної доброчесності визначені у розділі 3 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: https://kpi.ua/code.*

***Норми етичної поведінки.** Норми етичної поведінки студентів і працівників визначені у розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: https://kpi.ua/code.*

***Інклюзивне навчання.** Засвоєння знань та умінь в ході вивчення дисципліни «Сталий інноваційний розвиток» може бути доступним для більшості осіб з особливими освітніми потребами, окрім здобувачів з серйозними вадами зору, які не дозволяють виконувати завдання за допомогою персональних комп'ютерів, ноутбуків та/або інших технічних засобів.*

***Навчання іноземною мовою**. У ході виконання завдань студентам може бути рекомендовано звернутися до англомовних джерел.*

8. Види контролю та рейтингова система оцінювання результатів навчання (РСО)

Поточний контроль: модульна контрольна робота.

Модульна контрольна робота складається з двох частин – КР№1 і КР№2.

Кожна КР містить два завдання. Оцінки за теоретичні питання визначаються за шкалою:

- *«відмінно», повна відповідь (не менше 95% потрібної інформації) – 4.8-5 балів;*
- *«добре», достатньо повна відповідь (не менше 75% потрібної інформації), або повна відповідь з незначними неточностями – 3.7 – 4.8 балів;*
- *«задовільно», неповна відповідь (не менше 60% потрібної інформації) та значні помилки – 3 – 3.7 балів;*
- *«незадовільно», незадовільна відповідь (не відповідає вимогам на «задовільно») – 0 – 3 бали.*

Максимальна оцінка за кожну частину МКР складає 10 балів. ***Максимальна кількість балів за дві частини МКР складає 2*10 = 20 балів.***

***Календарний контроль:** проводиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу.*

***Семестровий контроль:** екзамен.*

Умови допуску до семестрового контролю: семестровий рейтинг не менше 40 балів.

Рейтингова система оцінювання результатів навчання

Рейтинг студента з кредитного модуля складається з балів, які він отримує за:

1) виконання та захист 12 (дванадцяти) лабораторних робіт – максимум 80 балів;
2) виконання модульної контрольної роботи – максимум 20 балів;
3) виконання творчих робіт за однією з тем дисципліни, опрацювання наукової літератури – додаткові 10 балів.

1. Практичні/ лабораторні роботи. Упродовж семестру студент має виконати 12 (дванадцять) практичних/ лабораторних робіт (ПР).

Рейтингова оцінка кожної ПР складається з 2 частин, які оцінюються окремо.

а. Якість підготовки до роботи, її виконання та оформлення звіту.

- за умови правильно оформленого звіту з точним виконанням завдання ПР – 3 – 3.5 бали;
- за наявності несуттєвих неточностей в оформленні або процедурі виконання ПР – 2 – 2.5 бали;
- за наявності порушень в оформленні, неповного або неточного виконання – 1 – 1.5 бали.

б. Якість захисту матеріалу. В цій частині оцінюється ступінь володіння теоретичним і практичним матеріалом за темою роботи.

- відмінне володіння матеріалом – 3 – 3.5 бали;
- добре володіння матеріалом – 2 – 2.5 бали;
- задовільне володіння матеріалом – 1 – 1.5 бали.

Максимальна кількість балів за всі ПР дорівнює **80 балів.**

2. Модульна контрольна робота. Модульна контрольна робота складається з двох частин – КР№1 і КР№2. Кожна КР містить два завдання. Оцінки за кожне завдання визначаються за шкалою:

- «відмінно», повна відповідь (не менше 95% потрібної інформації) – 4.8-5 балів;
- «добре», достатньо повна відповідь (не менше 75% потрібної інформації), або повна відповідь з незначними неточностями – 3.7 – 4.8 балів;
- «задовільно», неповна відповідь (не менше 60% потрібної інформації) та значні помилки – 3 – 3.7 балів;
- «незадовільно», незадовільна відповідь (не відповідає вимогам на «задовільно») – 0 – 3 бали.

Максимальна кількість балів за дві частини модульної КР складає 2*10 = **20 балів**.

3. Додаткові бали нараховуються за виконання творчих робіт за однією з тем дисципліни, опрацювання наукової літератури, до 10 балів.

За результатами навчальної роботи за перші 8 тижнів станом на 24.03 «ідеальний студент» має набрати 49 балів. На першому календарному контролі (8-й тиждень, 24.03) студент отримує «зараховано», якщо його поточний рейтинг не менше 49/2 = 25 балів.

За результатами 15 тижнів навчання станом на 12.05 «ідеальний студент» має набрати 100 балів. На другій атестації (15-й тиждень, 12.05) студент отримує «зараховано», якщо його поточний рейтинг не менше 60 балів.

Максимальна сума балів за роботу в семестрі складає 100. Необхідною умовою допуску до екзамену є отримання рейтингу 40 балів і вище. Для отримання екзамену з кредитного модуля «автоматом» потрібно мати рейтинг не менше 60 балів.

Студенти, які наприкінці семестру мають рейтинг менше 60 балів, а також ті, хто хоче підвищити оцінку, виконують екзаменаційну роботу. При цьому до балів за лабораторні роботи додаються бали за екзаменаційну роботу, і ця рейтингова оцінка є остаточною.

Завдання екзаменаційної контрольної роботи складається з чотирьох завдань різних розділів силабусу. Кожне завдання контрольної роботи оцінюється у 5 балів відповідно до системи оцінювання:

- «відмінно», повна відповідь (не менше 95% потрібної інформації) – 4.8-5 балів;
- «добре», достатньо повна відповідь (не менше 75% потрібної інформації), або повна відповідь з незначними неточностями – 3.7 – 4.8 балів;
- «задовільно», неповна відповідь (не менше 60% потрібної інформації) та значні помилки – 3 – 3.7 балів;
- «незадовільно», незадовільна відповідь (не відповідає вимогам на «задовільно») – 0 – 3 бали.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

Кількість балів	*Оцінка*
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Менше 40	Не допущено

9. Додаткова інформація з дисципліни (освітнього компонента)

Сертифікати проходження дистанційних чи онлайн курсів за тематикою дисципліни можуть бути зараховані з додатковими 5 – 10 балами до загального рейтингу студента.

Робочу програму навчальної дисципліни (силабус):

Складено професор, д.т.н., доц. Недашківська Надія Іванівна

Ухвалено кафедрою ММСА НН ІПСА (протокол № 13 від 05.06.2024)

Погоджено Методичною комісією НН ІПСА (протокол № 10 від 24.06.2024)