



Силабус освітнього компонента

Програма навчальної дисципліни



Некоректні задачі обробки даних

Шифр та назва спеціальності
113 – Прикладна математика

Інститут
ННІ Комп'ютерних наук та інформаційних технологій

Освітня програма
Інтелектуальний аналіз даних

Кафедра
Комп'ютерна математика і аналіз даних)

Рівень освіти
Магістр

Тип дисципліни
Спеціальна (фахова), Обов'язкова

Семестр
2

Мова викладання
Українська,

Викладачі, розробники



Любчик Леонід Михайлович

leonid.liubchik@khpi.edu.ua

Доктор технічних наук, професор

Кількість наукових та навчальних публікацій понад 200.

Провідний лектор з дисциплін: «Теорія керування», «Некоректні задачі обробки даних», «Прогнозний аналіз».

[Детальніше про викладача на сайті кафедри](http://web.kpi.kharkov.ua/kmmm/uk/o_kafedre_ua/profesorско-vikladatskij-sklad/lyubchik-leonid-mihajlovich/)

http://web.kpi.kharkov.ua/kmmm/uk/o_kafedre_ua/profesorско-vikladatskij-sklad/lyubchik-leonid-mihajlovich/



Ямковий Клим Сергійович

klym.yamkovyi@cs.khpi.edu.ua

асистент, PhD

Кількість наукових та навчальних публікацій 10.

[Детальніше про викладача на сайті кафедри](https://web.kpi.kharkov.ua/kmmm/uk/yamkovyj-klym-sergijovych-2/)

<https://web.kpi.kharkov.ua/kmmm/uk/yamkovyj-klym-sergijovych-2/>

Загальна інформація

Анотація

Дисципліна спрямована на вивчення методів вирішення обернених задач аналізу Великих Даних в умовах невизначеності. Розглядаються основні методи завдання моделей невизначеності, поняття і властивості обернених задач аналізу даних, а саме їх некоректності за Адамаром. Викладається основний метод розв'язання обернених некоректних задач великої розмірності в умовах невизначеності, а саме, метод регуляризації, розглядаються найбільш поширені методи регуляризації при статистичному і нестатистичному опису завад. Загальні методи ілюструються прикладами розв'язання типових некоректних задач обробки Великих Даних, оцінювання стану багатовимірних статичних і динамічних систем.

Мета та цілі дисципліни

Метою вивчення дисципліни є вивчення теорії і методів вирішення обернених некоректних задач аналізу невизначених Великих Даних, опанування методами регуляризації некоректних задач, методами оцінювання параметрів і стану статичних і динамічних систем великої розмірності за непрямими вимірюваннями у умовах невизначеності

Формат занять

Лекції, лабораторні роботи, розрахункові завдання, самостійна робота, консультації. Підсумковий контроль – іспит.

Компетентності

ЗК 7. Здатність працювати з інформацією: знаходити і використовувати інформацію з різних джерел, потрібну для розв'язання професійних завдань.

СК 1. Здатність формулювати математичну постановку задачі, спираючись на постановку мовою предметної галузі, перевіряти коректність постановки в умовах невизначеності.

СК 2 Здатність обирати, розробляти та досліджувати математичний аналітичний або чисельний метод розв'язання практичних задач, що забезпечує потрібні точність і надійність результату.

СК 5 Здатність до проведення математичного і комп'ютерного моделювання та обчислювального експерименту, збору, візуалізації, аналізу та обробки отриманих даних, розв'язання формалізованих задач за допомогою спеціалізованих програмних засобів.

СК 10. Здатність обирати, розробляти, досліджувати та застосовувати математичні моделі та методи для інтелектуального аналізу даних в умовах невизначеності.

СК 13 Здатність до розробки та експлуатації спеціалізованих програмних засобів обробки великих масивів даних на основі інформаційних технологій розподілених і хмарних обчислень.

СК 14 Здатність до використання сучасних інформаційних технологій інтелектуального аналізу даних, прогнозування, прийняття рішень, інформаційного пошуку і видобування знань.

Результати навчання

РН 2 Уміти формалізувати задачі, сформульовані мовою певної предметної галузі й обирати раціональний метод вирішення; розв'язувати задачі аналітичними або чисельними методами, оцінювати точність і достовірність отриманих результатів та виконувати їхню інтерпретацію.

РН 5 Будувати ефективні щодо точності обчислень, стійкості, швидкодії і витрат системних та обчислювальних ресурсів алгоритми для чисельного дослідження математичних моделей і аналізу даних, прийняття рішень.

РН 8 Уміти застосовувати у практичній роботі спеціалізовані програмні продукти і програмні системи комп'ютерної математики, аналізу великих даних тощо.

РН 12. Знати та розуміти сучасні методи розв'язання математичних задач статистичного і інтелектуального аналізу даних, прогнозування тощо.

РН 13 Знати і розуміти методи розв'язання математичних задач інтелектуального інформаційного пошуку та видобування знань.

РН 16 Уміти застосовувати сучасні інформаційні технології та програмне забезпечення для обробки великих масивів даних на основі розподілених і хмарних сервісів.

Обсяг дисципліни

Загальний обсяг дисципліни 150 год. (5 кредитів ECTS): лекції – 32 год., лабораторні роботи – 32 год., самостійна робота – 86 год.

Передумови вивчення дисципліни (пререквізити)

Для успішного проходження курсу необхідно мати знання та практичні навички з дисциплін: «Математичні методи машинного навчання 1», «Нелінійні процеси і моделі».

Особливості дисципліни, методи та технології навчання

Лекції проводяться інтерактивно з використанням мультимедійних технологій. Лабораторні роботи проводяться з використанням вільного програмного забезпечення – бібліотек SciPy, SkyLab, бібліотеки машинного навчання Scikit-Learn, екосистеми Apache Hadoop та її компонент.

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Оцінювання в умовах невизначеності.

Невизначені дані вимірювань і спостережень. Математичні моделі невизначених даних - інтервальних, статистичних, нечітких даних. Статистична та робастна оцінка.

Тема 2. Оборнені задачі аналізу великих даних

Проблеми аналітики великих даних. Відновлення функцій і полів, обробка сигналів і зображень. Оборнені задачі великої розмірності.

Тема 3. Некоректні задачі обробки даних

Коректність за Адамаром. Чутливість рішення до шуму.

Погано-обумовлені системи лінійних матричних рівнянь. Число обумовленості матриці.

Сингулярний розклад (SVD). Власні та сингулярні числа та власні вектори.

Сингулярний розклад квадратної та прямокутної матриць.

Чисельні методи SVD матриць великої розмірності.

Тема 4. Регуляризація некоректних задач

Регуляризація за А. Н. Тихоновим. Вибір параметра регуляризації.

Регуляризація а допомогою усіченого SVD.

Задача відновлення регресії, мультиколінеарність.

Регуляризована регресія. Регресія хребта та ласо.

Регресія в умовах невизначеності. Класи розподілів як моделі невизначених даних.

Вибір функції втрат, функція втрат Вапника-Гюбера.

Тема 5. Регуляризація в умовах невизначеності.

Регуляризація при статистичному описі невизначених даних.

Регуляризація при інтервальному описі невизначених даних.

Тема 6. Оцінка стану та входу багатовимірних системи.

Статистичне оцінювання стану статичних систем великої розмірності.

Рекурентне статистичне оцінювання стану багатовимірних динамічних систем.

Схема «прогноз - відновлення», фільтр Калмана.

Відновлення вхідних сигналів динамічних систем.

Оборнені динамічні моделі. Синтез на основі теорії спостерігачів для невідомого вхідного сигналу.

Теми практичних занять

Практичні заняття в рамках дисципліни не передбачені.

Теми лабораторних робіт

Тема 1. Статистична оцінка невизначених даних

Лаб. 1. Дослідження моделей невизначених даних.

Лаб. 2. Дослідження статистичної та робастної оцінки.

Тема 2. Вивчення чисельних методів і алгоритмів

Лаб. 3. Вивчення чисельних методів розв'язання системи лінійних матричних рівнянь

Лаб. 4. Вивчення чисельних методів діагоналізації та сингулярного розкладання матриць.

Тема 3. Вивчення чисельних методів оптимізації

Лаб. 5. Дослідження чисельних методів мінімізації квадратичного функціоналу.

Лаб. 6. Вивчення чисельних методів розв'язування некоректної регресійної задачі відновлення.

Тема 4. Вивчення оцінки стану системи

Лаб. 7. Дослідження чисельних методів статистичного оцінювання стану статичних систем.

Лаб. 8. Дослідження чисельних методів статистичного оцінювання стану динамічних систем.

Тема 5. Вивчення інформаційних технологій обробки Великих Даних.

Лаб. 9. Вивчення екосистеми Hadoop для обробки Великих Даних.

Вивчення та застосування алгоритмів MapReduce та технологій розподілених сховищ даних.

Лаб. 10. Ознайомлення зі Spark. Використання Apache Spark mllib для обробки потоків даних.

Самостійна робота

Тема 1. Вивчення методів опису невизначеності даних. Вивчення методів надійної статистики. Робота над прикладами поганообумовлених систем. Вивчення теорії сингулярного розкладання.

Тема 2. Вивчення теорії оптимізації при обмеженнях типу нерівностей.

Метод множників Лагранжа, метод штрафних функцій.

Тема 3. Вивчення теорії регуляризації за методом А.Н. Тихонова.

Вивчення методу найменших квадратів. Вивчення теорії регуляризованих МНК.

Тема 4. Вивчення методів статистичної оцінки стану статичних систем.

Вивчення алгоритму фільтра Калмана. Вивчення теорії зворотних динамічних моделей.

Тема 5. Застосування алгоритмів MapReduce та технологій розподілені обробки даних.

Вивчення бібліотеки Java MapReduce.

Тема 6. Екосистема Hadoop для Великих Даних.

Компоненти екосистеми Hadoop для обробки Великих Даних.

Тема 7. Обробка Великих даних в реальному часі.

Вивчення та застосування Apache Spark, Kafka.

Література та навчальні матеріали

Основна література

1. Некоректно поставлені задачі та методи їх розв'язування: Підручник / М.Г. Охріменко, І.Д. Фартушний, А.Б. Кулик . К.: В-во «Політехніка», 2018.

<https://docplayer.net/124193410-Nekorektno-postavleni-zadachi-ta-metodi-yih-rozv-yazuvannya.html>

2. Л.М. Олещенко. Технології оброблення великих даних. Конспект лекцій. Київ КПІ ім. Ігоря Сікорського 2021. 227 с.

https://ela.kpi.ua/bitstream/123456789/42206/1/%D0%9AonspLekts%20Tekhnolohii-obroblennia-velykykh-danykh_%D0%9Eleshchenko.pdf

3. Методичні вказівки з навчальної дисципліни «Методи розв'язування некоректних задач» для спеціальності 113 «Прикладна математика» / Бомба А. Я., Бойчура М. В. Рівне : НУВГП, 2021.

<https://ep3.nuwm.edu.ua/21354/1/04-01-53%D0%9C.pdf>

4. Yury Korolev, Jonas Latz. Inverse Problems. Lecture notes. University of Cambridge, 2020.

[https://www.damtp.cam.ac.uk/research/cia/files/teaching/Inverse Problems 2020/LectureNotes2020.pdf](https://www.damtp.cam.ac.uk/research/cia/files/teaching/Inverse%20Problems%2020/LectureNotes2020.pdf)

5. Christian Clason. Regularization of inverse problems. Lecture notes. Graz Uni. 2020.

<https://imsc.uni-graz.at/clason/skripte/InverseNotes.pdf>

6. David Di Ruscio. State estimation and Kalman filter. University of South-Eastern Norway. 2019.

http://www.davidr.no/iaa2217/pensum/main_estim_e2.pdf

Додаткова література

1. Андруник В.А., Висоцька В.А., Пасічник В.В., Чирун Л.Б., Чирун Л.В. Чисельні методи в комп'ютерних науках: навчальний посібник – Львів: Видавництво «Новий світ – 2000», 2020.

2. Guillaume Bal. Introduction to Inverse Problems. University of Chicago, Chicago. 2019.

<https://www.stat.uchicago.edu/~guillemebal/PAPERS/IntroductionInverseProblems.pdf>

https://pdf.lib.vntu.edu.ua/books/2019/Andrunik_P1_2017_470.pdf

3. Lectures on Inverse Problems and Imaging. GitHub Pages. 2024.

https://tristanvanleeuwen.github.io/IP_and_Im_Lectures/intro.html

4. Introduction to Big Data Analytics. Publisher: The Palm Publications. 2022.

https://www.researchgate.net/publication/360410918_INTRODUCTION_TO_BIG_DATA_ANALYTICS

Електронні ресурси

1. Introduction to Scilab. Режим доступу [http://www.bad.org.tr/mate/Belgeler/Scilab Manua in.pdf](http://www.bad.org.tr/mate/Belgeler/Scilab_Manua_in.pdf)

<https://www.scilab.org/>

2. Machine-learning software Skikit. Режим доступу <https://scikit-learn.org/stable/>

3. Apache Hadoop. Режим доступу: <http://hadoop.apache.org/>

4. MapReduce Tutorial. Режим доступу: [https://hadoop.apache.org/docs/r1.2.1/mapred tutorial.html](https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html)

5. Apache Spark. Режим доступу: <https://spark.apache.org/>

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

Бали студента з дисципліни нараховуються за наступним співвідношенням: контрольні роботи: 20% семестрової оцінки; самостійна робота: 15% семестрової оцінки; колоквіум: 25% семестрової оцінки; іспит: 40% семестрової оцінки.

Шкала оцінювання

Сума балів	Національна оцінка	ECTS
90–100	Відмінно	A
82–89	Добре	B
75–81	Добре	C
64–74	Задовільно	D
60–63	Задовільно	E
35–59	Незадовільно (потрібне додаткове вивчення)	FX
1–34	Незадовільно (потрібне повторне вивчення)	F

Норми академічної етики і політика курсу

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при неможливості вирішення конфлікту – доводитися до відома співробітників дирекції інституту. Нормативно-правове забезпечення впровадження принципів академічної доброчесності НТУ «ХПІ» розміщено на сайті: <http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/>.

Погодження

Силабус погоджено

Дата погодження, підпис
31.08.2023 р.

Завідувач кафедри
Олена АХІЄЗЕР

Дата погодження, підпис
31.08.2023 р.

Гарант ОП
Олексій ГАЛУЗА