

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Механіко-математичний факультет
Кафедра теорії функцій і функціонального аналізу

Затверджено
на засіданні кафедри теорії функцій
і функціонального аналізу
механіко-математичного факультету
Львівського національного
університету імені Івана Франка
(протокол № 1 від 27.08. 2020 р.)

В.о. завідувача кафедри
проф. Олег СКАСКІВ



Силабус з навчальної дисципліни
“Машинне навчання 2”,
що викладається в межах ОПП “Статистичний аналіз даних”
першого (бакалаврського) рівня вищої освіти для здобувачів з
спеціальності 112 – статистика

Львів 2020 р.

Назва дисципліни	Машинне навчання 2
Адреса викладання дисципліни	Головний корпус ЛНУ ім. І. Франка м. Львів, вул. Університетська 1
Факультет та кафедра, за якою закріплена дисципліна	Механіко-математичний факультет Кафедра теорії функцій і функціонального аналізу
Галузь знань, шифр та назва спеціальності	11 — математика та статистика: 112 – статистика
Викладачі дисципліни	Християнин Андрій Ярославович, доцент кафедри теорії функцій і функціонального аналізу
Контактна інформація викладачів	andriy.khrystiyany@lnu.edu.ua Головний корпус ЛНУ ім. І. Франка, каб. 373. м. Львів, вул. Університетська, 1
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекцій/практичних занять (за попередньою домовленістю).
Інформація про дисципліну	Дисципліна “Машинне навчання 2” є нормативною дисципліною з спеціальності 112 – статистика для освітньої програми “Статистичний аналіз даних”, яка викладається в 5-му семестрі в обсязі 3 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Курс спрямований на оволодіння класичними методами машинного навчання, теоретичними положеннями та основними застосуваннями методів машинного навчання в різноманітних задачах статистичного аналізу даних.
Мета та цілі дисципліни	Оволодіти класичними методами машинного навчання, теоретичними положеннями та основними застосуваннями методів машинного навчання в різноманітних задачах статистичного аналізу даних.
Література для вивчення дисципліни	1. A. Geron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed., 2020 2. T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning , 2017

	<p>3. S. Marsland, Machine learning: an algorithmic perspective, 2015 4. P. Flach, Machine Learning, 2013 5. C. Bishop, Pattern recognition and machine learning, 2007</p>
Обсяг курсу	<p>Загальний обсяг: 90 годин. Аудиторних занять: 48 год., з них 32 годин лекцій та 16 годин практичних робіт. Самостійної роботи: 42 година</p>
Очікувані результати навчання	<p>Після завершення цього курсу студент буде :</p> <ul style="list-style-type: none"> - Знати: основні поняття машинного навчання, зокрема: функція втрат, функціонал якості, градієнт, методи навчання моделей, перенавчання, регуляризація, відбір ознак, метрики якості класифікації, софтмакс, аргмакс, крос-валідація, опорні вектори, логічні закономірності, ядро, дерева ухвалення рішень, жадібні стратегії побудови логічних класифікаторів, бегінг та бустинг, кластеризація, нейронні мережі <p>Вміти:</p> <ul style="list-style-type: none"> - розв'язувати задачі регресії та класифікації, застосовувати метод найменших квадратів, метод градієнтного спуску, виконувати регуляризацію моделі і обирати оптимальний параметр регуляризації, використовувати регуляризацію для відбору ознак, розпізнавати перенавчання та недонавчання моделі, боротися з перенавчанням, виконувати крос-валідацію, виконувати перетворення ознак, розв'язувати задачі класифікації та регресії за допомогою методу опорних векторів, застосовувати метод k найближчих сусідів та його узагальнення, будувати логічні класифікатори, застосовувати метод випадкового лісу для задач класифікації та регресії, застосовувати методи бегінгу, бустингу для побудови композицій базових алгоритмів для розв'язування задач класифікації та регресії, застосовувати методи k-means та DBSCAN для розв'язування задач кластеризації, застосовувати ітеративні алгоритми Expectation-Maximization для відновлення щільності розподілу, конфігурувати та навчати нейронні мережі для розв'язування задач класифікації та регресії. - використовувати бібліотеки scikit-learn, numpy, pandas для розв'язування задач машинного навчання <p>В результаті засвоєння матеріалу даного курсу студент набуде таких загальних (ЗК) і спеціальних (фахових) (СК) компетентностей:</p> <p>ЗК-1. Здатність до абстрактного мислення, аналізу та синтезу. ЗК-2. Здатність застосовувати знання у практичних ситуаціях. ЗК-3. Знання й розуміння предметної області та професійної</p>

діяльності.

ЗК-5. Здатність спілкуватися іноземною мовою.

ЗК-6. Навички використання інформаційних і комунікаційних технологій.

ЗК-7. Здатність вчитися і оволодівати сучасними знаннями.

ЗК-8. Здатність до пошуку, обробки та аналізу інформації з різних джерел.

ЗК-10. Здатність працювати в команді.

ЗК-11. Здатність до професійного спілкування з представниками інших професійних груп різного рівня (з експертами в інших галузях знань).

ЗК-12. Здатність працювати автономно.

ЗК-13. Визначеність і наполегливість щодо поставлених завдань і взятих обов'язків.

СК-2. Здатність застосовувати у професійній діяльності знання та навички в галузях теорії ймовірностей, математичної статистики, теорії випадкових процесів.

СК-3. Здатність здійснювати логічні математичні міркування із чітким зазначенням припущень та висновків.

СК-4. Здатність до математичного формулювання задач та вибору методів їх розв'язання.

СК-5. Здатність до кількісно-статистичного мислення.

СК-6. Здатність до ймовірнісного мислення, що передбачає сприйняття стохастичної природи явищ.

СК-7. Здатність робити якісні висновки з кількісних даних.

СК-8. Уміння працювати з інформаційними базами даних.

СК-9. Здатність розробляти експериментальні та спостережувальні дослідження та аналізувати дані цих досліджень.

СК-10. Здатність проводити дослідження ймовірнісно-статистичних моделей та інтерпретувати одержані результати.

СК-11. Здатність використання обчислювальної техніки, спеціалізованих мов програмування та програмних засобів для розв'язання задач і здобуття додаткової інформації.

СК-12. Здатність застосовувати ймовірнісно-статистичні методи в міждисциплінарному контексті.

СК-13. Здатність подавати статистичні процедури та результати їхнього застосування у формі, придатній для цільової аудиторії, до якої звертаються, як усно, так і письмово.

СК-14. Здатність до аналізу основ і властивостей статистичних алгоритмів та розуміння переваг і обмежень тих чи інших підходів, у тому числі до оцінки їх обґрунтованості й ефективності.

СК-15. Здатність аналізувати основи і властивості базових економічних та фінансових структур, інтерпретувати показники фінансової діяльності, користуватися методами оптимального

	<p>керування економічних та природних процесів.</p> <p>СК-16. Здатність застосовувати у професійній діяльності знання та навички з машинного навчання, обробки зображень і природної мови.</p> <p>СК-17. Здатність моделювати та пояснювати дані просторових і часових вибірок за допомогою знань і навичок з регресійного аналізу.</p> <p>і здобуде такі результати навчання (РН):</p> <p>РН-2. Вміти працювати зі спеціальною літературою іноземною мовою.</p> <p>РН-11. Вміти аналізувати та прогнозувати лінійні статистичні моделі та моделі регресії, оцінювати їхні параметри.</p> <p>РН-12. Вміти збирати та обробляти дані, застосовувати статистичні процедури для аналізу даних за допомогою обчислювальної техніки та програмних засобів.</p> <p>РН-14. Володіти сучасними інформаційними технологіями для створення презентацій, роботи з базами даних, пошуку інформації та обміну нею.</p> <p>РН-17. Знати методи моделювання природних та/або соціальних процесів.</p> <p>РН-18. Вміти застосовувати ймовірнісно-статистичні моделі та методи для розв'язання прикладних проблем і задач.</p> <p>РН-19. Вміти оцінювати та забезпечувати якість виконуваних досліджень.</p> <p>РН-20. Вміти використовувати ймовірнісні та статистичні методи аналізу фінансових показників діяльності суб'єктів ринку.</p> <p>РН-21. Вміти застосовувати у професійній діяльності знання та навички з машинного навчання, обробки зображень і природної мови, інших галузей науки про дані.</p>
<p>Ключові слова</p>	<p>лінійна регресія, лінійна класифікація, логістична регресія, метод k найближчих сусідів, метод опорних векторів, сингулярний розклад, метод головних компонент, дерева ухвалення рішень, випадковий ліс, об'єкти, ознаки, функція втрат, функціонал якості, метод градієнтного спуску, регуляризація, відбір ознак, кластеризація, нейронні мережі.</p>
<p>Формат курсу</p>	<p>Очний.</p> <p>Проведення лекцій, практичних занять і консультацій.</p>
<p>Теми</p>	<ol style="list-style-type: none"> 1. Композиції алгоритмів (ансамблі). Агрегуючі функції та вирішальні правила. 2. Проблема різноманітності базових алгоритмів. 3. Стохастичні ансамблеві методи: bagging, pasting, random subspaces method, random patches. 4. Незміщена оцінка помилок. Оцінювання важливості ознак. 5. Бустинг: AdaBoost, основна теорема та теорема про збіжність. 6. Порівняння бустингу та бегінгу.

	<p>7. ComBoost. 8. Градієнтний бустинг: XGBoost. 9. Нелінійні ансамблеві методи: блендінг, стекінг. 10. Квазілінійні композиції алгоритмів: функції компетентності. 11. Задача відшукування нетипових об'єктів. 12. Опуклі функції втрат. 13. Алгоритм Mixture of Experts. 14. Аналіз bias/variance. 15. Кластеризація та часткове навчання. 16. Критерії якості кластеризації. 17. Графові алгоритми кластеризації. 18. Алгоритм k-means. 19. Алгоритм DBSCAN. 20. Ієрархічні методи кластеризації. Алгоритм Ланса-Вільямса. 21. Часткове навчання: алгоритми self-training, co-training, co-learning, трансдуктивний SVM. 22. Логістична регресія та часткове навчання : алгоритм Expectation-Regularization. 23. Методи відновлення щільності розподілу. Оптимальний байєсівський класифікатор. 24. Методи відновлення щільності розподілу: непараметричний та параметричний підхід. 25. Алгоритм Expectation-Minimization. GaussianMixtureModels. 26. Штучні нейронні мережі. 27. Метод зворотнього поширення помилок. 28. Dropout, функції активації, BatchNorm 29. Ініціалізація ваг. 30. Проріджування ШНН.</p>
Підсумковий контроль, форма	Екзамен
Пререквізити	Для вивчення курсу студенти потребують знань курсів: машинне навчання 1, математичного аналізу, лінійної алгебри, дискретної математики, основ теорії ймовірностей та статистики, основ програмування мовою python
Навчальні методи та техніки, які будуть використовуватися під час викладання	Презентації, лекції, індивідуальні завдання.

курсу	
Необхідне обладнання	Аудиторія обладнана дошкою та засобами написання для аудиторних занять, комп'ютер із відповідним програмним забезпеченням для виконання практичних завдань.
Критерії оцінювання (окремо для кожного виду навчальної діяльності)	<p>Оцінювання проводиться за 100-бальною шкалою. Бали нараховуються за наступним співвідношенням:</p> <p>теоретичні колоквиуми : 20% семестрової оцінки індивідуальні завдання : 30% семестрової оцінки максимальна кількість балів протягом семестру : 50 екзамен: 50% семестрової оцінки; максимальна кількість балів на екзамені : 50 Підсумкова максимальна кількість балів : 100.</p> <p>Академічна доброчесність: Очікується, що роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикування джерел, списування, втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в письмовій роботі студента є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману.</p> <p>Відвідування занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції та практичні заняття курсу. Студенти повинні інформувати викладача про неможливість відвідати заняття. У будь-якому випадку студенти зобов'язані дотримуватися термінів визначених для виконання всіх видів письмових робіт та індивідуальних завдань, передбачених курсом.</p> <p>Література. Уся література, яку студенти не зможуть знайти самостійно, буде надана викладачем виключно в освітніх цілях без права її передачі третім особам. Студенти заохочуються до використання також й іншої літератури та джерел, яких немає серед рекомендованих.</p> <p>Політика виставлення балів. Враховуються бали набрані при поточному тестуванні та бали підсумкового тестування. При цьому обов'язково враховуються присутність на заняттях та активність студента під час практичного заняття; недопустимість пропусків та запізнь на заняття; користування мобільним телефоном, планшетом чи іншими мобільними пристроями під час заняття в цілях не пов'язаних з навчанням; списування та плагіат; несвоєчасне виконання поставленого завдання і т. ін.</p> <p>Жодні форми порушення академічної доброчесності не толеруються.</p> <p>Оцінювання практичних робіт. Два індивідуальні завдання протягом семестру по 15 балів кожне. Індивідуальні завдання</p>

	<p>ґрунтуються на матеріалах практичних занять. Виконання завдань передбачає використання програмних засобів використовуваних в межах курсу. Бали нараховуються за виконання окремих пунктів кожного індивідуального завдання.</p> <p>Оцінювання колоквиуму. Колоквиум проводиться у формі тестування з використанням університетського серверу e-learning.lnu.edu.ua. Максимальна кількість балів : 20.</p> <p>Відсотки нарахування балів оцінювання відповіді на кожне запитання нараховуються за наступним співвідношенням:</p> <p>75-100% – тема відтворюється в повному обсязі, правильно, обґрунтовано, логічно;</p> <p>50-75% – відтворюється значна частина розглянутої теми, проте присутні неточності та/або невідповідності;</p> <p>25-50% – виявлено множинні неточності та невідповідності, пояснення відсутні чи частково помилкові;</p> <p>0-25% – тему майже не розкрито, кількість викладеного матеріалу не відповідає загальним нормам обраного виду роботи.</p> <p>Критерії оцінювання результатів неформальної освіти. Нарахування балів відбувається за публікацію студентом тез доповідей на конференціях, наукових статей, за участь студента у діяльності наукових гуртків, семінарів, круглих столів, конкурсів, участь у заходах неформальної освіти, за отримання сертифікатів про проходження навчання на різних освітніх платформах (Coursera, Prometheus тощо), курсах провідних ІТ компаній за тематикою навчальної дисципліни. Кількість балів визначається відсотком покриття результатів відповідної активності до вимог результатів навчання з навчальної дисципліни.</p>
Питання до екзамену.	На екзамені виносяться усі теми курсу (див. пункт Теми).
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

Схема курсу

Т и ж н і	Лекційний курс		Практичні заняття		К - с т ь г о д С Р	Літера- тура
	Назва теми	К - с т ь г о д	Назва теми	К - с т ь г о д		
1	2	3	4	5	6	7
1	Композиції алгоритмів. Агрегуючі функції. Вирішальні правила. Проблема різноманітності базових алгоритмів. (Теми 1-2)	2	Композиції алгоритмів.	2	2	[1], [2], [3], [4], [5]
2	Стохастичні ансамблеві методи. Незміщена оцінка помилок. Оцінювання важливості ознак. (Теми 3-4)	2		0	3	[1], [2], [3], [4], [5]
3	Бустинг: AdaBoost. Основна теорема бустингу. Збіжність процесу навчання. (Тема 5)	2	Стохастичні ансамблеві методи засобами python, numpy, sklearn. Бустинг	2	2	[1], [2], [3], [4], [5]
4	Оцінка ймовірності помилок композиції. ComBoost. Керування розподілом відступів для підвищення різноманітності базових алгоритмів. (Теми 6-7)	2		0	3	[1], [2], [3], [4], [5]
5	Гradientний бустинг. Стохастичний gradientний бустинг. XGBoost. (Тема 8)	2	Gradientний бустинг засобами python, numpy, sklearn. Нелінійні композиції алгоритмів	2	2	[1], [2], [3], [4], [5]
6	Нелінійні композиції алгоритмів: блендінг і стекінг. Квазілінійні композиції: MixtureOfExperts. Використання опуклих функцій втрат. (Теми 9-13)	2		0	3	[1], [2], [3], [4], [5]
7	Аналіз зсуву/дисперсії. (Тема 14)	2	Захист Індивідуального завдання № 1	2	2	[1], [2], [3], [4], [5]
8	Кластеризація. Критерії якості кластеризації. Графові алгоритми кластеризації. (Теми 15-17)	2		0	3	[1], [2], [3], [4], [5]
9	Алгоритми кластеризації k-means та DBSCAN. (Теми 18-19)	2	Кластеризація засобами python, sklearn. k-means за допомогою python, sklearn	2	2	[1], [2], [3], [4], [5]
10	Ієрархічні методи кластеризації. Алгоритм Ланса-Вільямса. Дендрограми. (Теми 20)	2		0	3	[1], [2], [3], [4], [5]
11	Часткове навчання. Self-training, co-learning,	2	DBSCAN за допомогою python, sklearn. Ієрархічні методи	2	2	[1], [2], [3], [4], [5]

	трансдуктивний SVM, регуляризація правдоподібності. (Теми 21- 22) Колоквіум		кластеризації. Часткове навчання (засобами python, scipy, sklearn)			
12	Відновлення щільності розподілу. Оптимальний баєсівський класифікатор. NaiveBayes (Тема 23)	2		0	3	[1], [2], [3], [4], [5]
13	Відновлення одновимірної щільності розподілу. Непараметричний та параметричний підходи. (Тема 24)	2	Захист Індивідуального завдання № 2	2	3	[1], [2], [3], [4], [5]
14	Відновлення суміші розподілів. EM-алгоритм. Застосування до кластеризації. Gaussian Mixture Models. (Тема 25)	2		0	3	[1], [2], [3], [4], [5]
15	Штучні нейронні мережі. Метод зворотнього поширення помилки. (Теми 26-27)	2	Відновлення щільності розподілу засобами python, numpy, scipy. GMM та EM засобами python, numpy, sklearn. Штучні нейронні мережі	2	3	[1], [2], [3], [4], [5]
16	Штучні нейронні мережі. Покращення збіжності. Функції активації. Регуляризація. Ініціалізація ваг. (Теми 28-30)	2		0	3	[1], [2], [3], [4], [5]
	Разом	32		16	42	
	Викладач: Християнин А. Я.		Викладач: Християнин А. Я.			