

**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 1”,**  
**що викладається в межах ОПП “Статистичний аналіз даних”**  
**першого (бакалаврського) рівня вищої освіти для здобувачів з**  
**спеціальності 112 – статистика**

**Львів 2020 р.**

<b>Назва дисципліни</b>	Машинне навчання 1
<b>Адреса викладання дисципліни</b>	Головний корпус ЛНУ ім. І. Франка м. Львів, вул. Університетська 1, 79000
<b>Факультет та кафедра, за якою закріплена дисципліна</b>	Механіко-математичний факультет Кафедра теорії функцій і функціонального аналізу
<b>Галузь знань, шифр та назва спеціальності</b>	11 — математика та статистика: 112 – статистика
<b>Викладачі дисципліни</b>	Християнин Андрій Ярославович, кандидат фізико-математичних наук, доцент кафедри теорії функцій і функціонального аналізу
<b>Контактна інформація викладачів</b>	<a href="mailto:andriy.khrystiyany@lnu.edu.ua">andriy.khrystiyany@lnu.edu.ua</a>  Головний корпус ЛНУ ім. І. Франка, каб. 373. м. Львів, вул. Університетська, 1
<b>Консультації з питань навчання по дисципліні відбуваються</b>	Консультації в день проведення лекцій/практичних занять (за попередньою домовленістю).
<b>Інформація про дисципліну</b>	Дисципліна “Машинне навчання 1” є нормативною дисципліною з спеціальності 112 – статистика для освітньої програми Математика, яка викладається в 4-му семестрі в обсязі 3,5 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
<b>Коротка анотація дисципліни</b>	Курс спрямований на оволодіння класичними методами машинного навчання, теоретичними положеннями та основними застосуваннями методів машинного навчання в різноманітних задачах статистичного аналізу даних.
<b>Мета та цілі дисципліни</b>	Оволодіти класичними методами машинного навчання, теоретичними положеннями та основними застосуваннями методів машинного навчання в різноманітних задачах статистичного аналізу даних.
<b>Література</b>	1. A. Geron, Hands-On Machine Learning with Scikit-Learn, Keras,

<p><b>для вивчення дисципліни</b></p>	<p>and TensorFlow, 2nd ed., 2020  2. T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning , 2017  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>
<p><b>Обсяг курсу</b></p>	<p>Загальний обсяг: 105 годин. Аудиторних занять: 64 год., з них 32 годин лекцій та 32 годин лабораторних робіт. Самостійної роботи: 41 година</p>
<p><b>Очікувані результати навчання</b></p>	<p>Після завершення цього курсу студент буде :</p> <p><b>Знати:</b>  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</p> <p><b>Вміти:</b>  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.  використовувати бібліотеки scikit-learn, numpy, pandas для розв'язування задач машинного навчання.</p> <p>В результаті засвоєння матеріалу даного курсу студент набуде таких <b>загальних (ЗК) і спеціальних (фахових) (СК) компетентностей:</b>  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 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.</p>

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 керування економічних та природних процесів.

СК-16. Здатність застосовувати у професійній діяльності знання та навички з машинного навчання, обробки зображень і природної мови.

СК-17. Здатність моделювати та пояснювати дані просторових і часових вибірок за допомогою знань і навичок з регресійного аналізу.

	<p>і здобуде такі <b>результати навчання (РН)</b>:</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 природничих 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><b>Ключові слова</b></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.</p>
<p><b>Формат курсу</b></p>	<p>Очний Проведення лекцій, практичних занять і консультацій.</p>
<p><b>Теми</b></p>	<ol style="list-style-type: none"> <li>1. Постановка задачі навчання за прецедентами.</li> <li>2. Об'єкти та ознаки.</li> <li>3. Модель алгоритмів та метод навчання.</li> <li>4. Функція втрат та функціонал якості.</li> <li>5. Зведення задачі навчання до задачі оптимізації.</li> <li>6. Ймовірнісна постановка задачі навчання.</li> <li>7. Принцип максимуму правдоподібності.</li> <li>8. Проблеми перенавчання та недонавчання.</li> <li>9. Узагальнююча здатність методу навчання.</li> <li>10. Лінійні моделі.</li> <li>11. Використання категорійних та текстових ознак.</li> <li>12. Бінарізація ознак та one-hot кодування.</li> <li>13. Лінійна регресія.</li> <li>14. Явний розв'язок задачі відновлення лінійної регресії.</li> </ol>

15. Градієнт. Метод градієнтного спуску.
16. Метод стохастичного градієнтного спуску.
17. Рекурентна оцінка функціоналу якості.
18. Метод накопичення імпульсу (momentum).
19. NAG (Nesterov Accelerated Gradient).
20. Варіанти ініціалізації ваг.
21. Вибір градієнтного кроку.
22. Методи другого порядку: метод Левенберга-Марквардта.
23. Регуляризація: L1 та L2-регуляризації.
24. Сингулярний розклад. Розв'язок задачі МНК через сингулярний розклад матриці об'єктів-ознак.
25. Регуляризований метод найменших квадратів через сингулярний розклад. Вибір параметра регуляризації.
26. L1-регуляризація як відбір ознак.
27. Метод головних компонент.
28. Лінійна класифікація. Порогова функція втрат.
29. Поняття відступу та дискримінантна функція.
30. Неперервні функції втрат.
31. Ймовірнісне формулювання задачі класифікації.
32. Максимізація правдоподібності та мінімізація емпіричного ризику.
33. Ймовірнісний зміст регуляризації. Розподіли Лапласа та Гаусса.
34. Принцип спільної правдоподібності даних і моделі.
35. Двокласова логістична регресія. Логарифмічна функція втрат.
36. Багатокласова логістична регресія.
37. Аналіз помилок класифікації. Точність і повнота.
38. Метрики якості бінарної класифікації.
39. ROC-крива. Площа під ROC-кривою як метрика якості.
40. Ймовірнісний зміст AUC-ROC.
41. Градієнтна максимізація AUC-ROC.
42. PR-крива. Площа під PR-кривою як метрика якості.
43. Точність і повнота для багатокласової класифікації.
44. Метод опорних векторів.
45. Умови Кароша-Куна-Такера. Розв'язок оптимізаційної задачі.
46. Метод опорних векторів: випадок лінійно нерозділюваної вибірки.
47. Типізація об'єктів. Опорні об'єкти.
48. Нелінійні узагальнення лінійних моделей. Спрямний простір.
49. Нелінійні узагальнення методу опорних векторів.
50. Ядра. Kernel trick. Методи генерації ядер.
51. SVM-регресія.
52. Подвійна регуляризація SVM (ElasticNet SVM)

	<p>53. Відбір ознак. Support Features Machines.</p> <p>54. Метричні методи класифікації та регресії.</p> <p>55. Метод k найближчих сусідів.</p> <p>56. Метод k зважених сусідів.</p> <p>57. Метод вікна Парзена. Зв'язок з лінійним класифікатором.</p> <p>58. Метод потенціальних функцій. Відбір еталонних об'єктів.</p> <p>59. Повна крос-валідація. Профіль компактності.</p> <p>60. Непараметрична регресія. Формула Надарая-Ватсона</p> <p>61. Вибір ядра та ширини вікна.</p> <p>62. Локально зважене згладжування.</p> <p>63. Логічні закономірності та правила.</p> <p>64. Логічні алгоритми класифікації.</p> <p>65. Поняття інформативності. Критерії інформативності.</p> <p>66. Ентропійний критерій інформативності. Приріст інформації.</p> <p>67. Багатокласова інформативність.</p> <p>68. Список ухвалення рішень. Жадібний алгоритм побудови списку ухвалення рішень.</p> <p>69. Критерії відбору правил.</p> <p>70. Дерева рішень. Класифікація об'єктів за допомогою дерева ухвалення рішень.</p> <p>71. Алгоритм побудови дерева рішень.</p> <p>72. Критерії галуження. Критерій Джині.</p> <p>73. Робота з неповними даними.</p> <p>74. Недоліки жадібної стратегії побудови дерев.</p> <p>75. Редукція дерев.</p> <p>76. CART: застосування дерев до розв'язування задач відновлення регресії.</p> <p>77. Композиції дерев. Метод випадкового лісу.</p>
<b>Підсумковий контроль, форма</b>	Екзамен
<b>Пререквізити</b>	Для вивчення курсу студенти потребують знань курсів: математичного аналізу, лінійної алгебри, дискретної математики, основ теорії ймовірностей та статистики, основ програмування мовою python
<b>Навчальні методи та техніки, які будуть використовуватися під час викладання</b>	Презентації, лекції, індивідуальні завдання.

<b>курсу</b>	
<b>Необхідне обладнання</b>	Аудиторія обладнана дошкою та засобами написання для аудиторних занять, комп'ютер із відповідним програмним забезпеченням для виконання практичних завдань.
<b>Критерії оцінювання (окремо для кожного виду навчальної діяльності)</b>	<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><b>Академічна доброчесність:</b> 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><b>Відвідування занять</b> 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><b>Література.</b> 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><b>Політика виставлення балів.</b> 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><b>Оцінювання практичних робіт.</b> Два індивідуальні завдання</p>



	<p>протягом семестру по 15 балів кожне. Індивідуальні завдання 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><b>Оцінювання колоквиуму.</b> Колоквиум проводиться у формі тестування з використанням університетського серверу e-learning.lnu.edu.ua. Максимальна кількість балів : 20.</p> <p><b>Критерії оцінювання результатів неформальної освіти.</b> 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>
<b>Питання до екзамену.</b>	На екзамені виносяться усі теми курсу (див. пункт Теми).
<b>Опитування</b>	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	Типи задач машинного навчання. Зведення до задач оптимізації. (Теми 1-9)	2	Огляд основних інструментів та програмних засобів, які використовуватимуться в курсі	2	2	[1], [2], [3], [4], [5]
2	Лінійна регресія. (Теми 10-14)	2	Векторно-матричне диференціювання.	2	3	[1], [2], [3], [4], [5]
3	Метод градієнтного спуску. (Теми 15-17)	2	Явний розв'язок задачі відновлення лінійної регресії.	2	2	[1], [2], [3], [4], [5]

4	Узагальнення методу градієнтного спуску. Регуляризація. (Теми 18-23)	2	Лінійна регресія та метод градієнтного спуску засобами python, numpy, sklearn	2	3	[1], [2], [3], [4], [5]
5	Сингулярний розклад матриці. (Теми 24-25)	2	Сингулярний розклад матриці : аналітичний підхід.	2	2	[1], [2], [3], [4], [5]
6	Розв'язок задачі відновлення лінійної регресії за допомогою сингулярного розкладу. Метод головних компонент. (Теми 26-27)	2	Сингулярний розклад матриці та метод головних компонент засобами python, numpy, sklearn	2	3	[1], [2], [3], [4], [5]
7	Лінійна класифікація. (Теми 28-30)	2	Лінійна класифікація.	2	2	[1], [2], [3], [4], [5]
8	Логістична регресія. (Теми 31-36)	2	Логістична регресія.  Захист Індивідуального завдання №1	2	3	[1], [2], [3], [4], [5]
9	Метрики якості класифікації. (Теми 37-38)	2	Метрики якості класифікації.	2	2	[1], [2], [3], [4], [5]
10	Метрики якості класифікації. (Теми 39-43) Колоквіум.	2	Оцінювання метрик якості класифікації за допомогою sklearn	2	3	[1], [2], [3], [4], [5]
11	Метод опорних векторів. Випадок лінійно розділюваної вибірки. (Теми 44-45)	2	Метод опорних векторів : аналітичних підхід.	2	2	[1], [2], [3], [4], [5]
12	Метод опорних векторів. Випадок лінійно нерозділюваної вибірки. Узагальнення. (Теми 46-53)	2	Метод опорних векторів за допомогою sklearn  Захист Індивідуального завдання № 2	2	3	[1], [2], [3], [4], [5]
13	Метричні методи розв'язування задач класифікації та регресії. (Теми 54-57)	2	Метричні методи розв'язування задач класифікації та регресії за допомогою python, numpy, sklearn	2	2	[1], [2], [3], [4], [5]
14	Метричні методи розв'язування задач класифікації та регресії. (Теми 58-62)	2	Метричні методи розв'язування задач класифікації та регресії за допомогою python, numpy, sklearn	2	3	[1], [2], [3], [4], [5]
15	Логічні методи розв'язування задач класифікації та регресії. (Теми 63-68)	2	Критерії інформативності.	2	3	[1], [2], [3], [4], [5]
16	Логічні методи розв'язування задач класифікації та регресії. (Теми 69-77)	2	Логічні методи розв'язування задач класифікації та регресії за допомогою sklearn	2	3	[1], [2], [3], [4], [5]
	<b>Разом</b>	<b>32</b>		<b>32</b>	<b>41</b>	
	Викладач: <b>Християнин А. Я.</b>		Викладач: <b>Християнин А. Я.</b>			