

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Факультет прикладної математики та інформатики
Кафедра обчислювальної математики

Затверджено

на засіданні
кафедри обчислювальної математики
факультету прикладної математики та
інформатики
Львівського національного університету
імені Івана Франка
(протокол № 1 від 29 серпня 2023 р.)

Завідувач кафедри



Роман ХАПКО

Силабус з навчальної дисципліни
«Вступ до машинного навчання»,
що викладається в межах першого (бакалаврського) рівня
вищої освіти для здобувачів
зі спеціальності 113 Прикладна математика

Львів 2023 р.

Назва дисципліни	Вступ до машинного навчання
Адреса викладання дисципліни	Головний корпус ЛНУ ім. І. Франка м. Львів, вул. Університетська 1
Факультет та кафедра, за якою закріплена дисципліна	Факультет прикладної математики та інформатики Кафедра обчислювальної математики
Галузь знань, шифр та назва спеціальності	113 – прикладна математика
Викладачі дисципліни	Музичук Юрій Анатолійович, доцент кафедри обчислювальної математики, Борачок Ігор Володимирович, асистент кафедри обчислювальної математики
Контактна інформація викладачів	Yuriy.Muzychuk@lnu.edu.ua ; https://ami.lnu.edu.ua/employee/muzychuk-yuriy Ihor.Borachok@lnu.edu.ua ; https://ami.lnu.edu.ua/employee/borachok-ihor Головний корпус ЛНУ ім. І. Франка, каб. 262. м. Львів, вул. Університетська, 1
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекцій/лабораторних занять (за попередньою домовленістю).
Сторінка курсу	https://ami.lnu.edu.ua/course/osnovy-mashynnoho-navchannia
Інформація про дисципліну	Дисципліна «Вступ до машинного навчання» є дисципліною на вибір зі спеціальності 113 Прикладна математика, яка викладається в 6-му семестрі (5 кредитів ECTS).
Коротка анотація дисципліни	Курс «Вступ до машинного навчання» охоплює такі розділи: задачі регресії, задачі класифікації та сучасні проблеми машинного навчання. При розгляді цих тем основна увага зосереджується на математичному формулюванні задач, виборі методики розв'язування та детальному описі та програмуванні відповідних алгоритмів. Викладення матеріалу здійснюється за допомогою сучасних термінів та понять з галузі інформаційних технологій. Метою курсу є ґрунтовне ознайомлення із наявними методами розв'язування різноманітних задач з області машинного навчання та роз'яснення вагомих аспектів їх використання.
Мета та цілі дисципліни	Головним завданням курсу є ґрунтовне ознайомлення студентів з існуючими алгоритмами машинного навчання та існуючими сучасними технологіями для їхньої реалізації.
Література для вивчення дисципліни	<p style="text-align: center;">Основна література</p> <ol style="list-style-type: none"> 1. Goodfellow I. Deep Learning / Ian Goodfellow, Yoshua Bengio, Aaron Courville. – MIT Press, 2016. – WWW: http://www.deeplearningbook.org. 2. Witten I. Data Mining: Practical Machine Learning Tools and Techniques / Ian Witten, Eibe Frank, Mark A. Hall. – Morgan Kaufmann, 2016. – 654p. 3. Flach P. Machine Learning: The Art and Science of Algorithms that Make Sense of Data / Peter Flach. – Cambridge University Press, 2012. – 409p. 4. Raschka S. Python Machine Learning / Sebastian Raschka. – Packt Publishing, 2015. – 454p. 5. Géron A. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Techniques and Tools to Build Learning Machines / Aurélien Géron. – O'Reilly Media, 2018. – 566p.

	<p style="text-align: center;">Додаткова література</p> <ol style="list-style-type: none"> 6. Patterson J. Deep Learning: A Practitioner's Approach / Josh Patterson, Adam Gibson. – O'Reilly, 2017. – 352 p. 7. Burns S. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn and Tensorflow / Samuel Burns. – 2019. – 176 p. 8. Papa J. PyTorch Pocket Reference: Building and Deploying Deep Learning Models / Joe Papa. – O'Reilly, 2021. – 307 p. 9. Hope T. Learning TensorFlow: A Guide to Building Deep Learning Systems / Tom Hope, Yehezkel Resheff, Itay Lieder. – O'Reilly, 2017. – 242 p. 10. Foster D. Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play / David Foster. – O'Reilly, 2019. – 330 p. 11. Harrison P. Deep Learning with Text / Patrick Harrison, Matthew Honnibal. – O'Reilly, 2020. – 250 p.
Обсяг курсу	Загальний обсяг: 150 годин. Аудиторних занять: 64 год., з них 32 год. Лекцій та 32 години лабораторних робіт. Самостійної роботи: 86 год.
Очікувані результати навчання	<p>В результаті вивчення даного курсу студент повинен:</p> <p>Знати:</p> <ul style="list-style-type: none"> - математичне формулювання поширених сучасних задач машинного навчання, - найбільш поширені алгоритми машинного навчання; <p>Вміти:</p> <ul style="list-style-type: none"> - на основі поставлених вимог спроектувати та реалізувати нейронну мережу з використанням актуальної бібліотеки машинного навчання, - оцінити результати роботи побудованої мережі та запропонувати ідеї для покращення точності отриманих результатів. <p>Даний курс сприятиме формуванню та поглибленню таких соціальних, «м'яких» навичок (soft skills):</p> <ul style="list-style-type: none"> - <i>критичне мислення</i> – аналіз даних, розуміння складних алгоритмів та оцінка моделей машинного навчання стимулюють до аналітичного підходу до розв'язування поставлених задач. - <i>комунікація</i> – необхідність презентувати результати своїх індивідуальних проектів та пояснювати тонкощі роботи алгоритмів вдосконалюють здатність чітко та ефективно доносити до аудиторії складну інформацію. - <i>прийняття рішень</i> – вибір необхідних алгоритмів, наборів даних та параметрів для ефективного розв'язування поставлених задач спонукатиме до прийняття проінформованих рішень, орієнтованих на факти. - <i>гнучкість</i> – машинне навчання, як сфера комп'ютерних наук, яка активно розвивається через швидку появу нових технологій та методик, вимагають гнучкості мислення та постійного навчання.
Ключові слова	Машинне навчання, нейронна мережа, логістична регресія, лінійна регресія, метод градієнтного спуску.
Формат курсу	Очний Проведення лекцій, лабораторних робіт і консультацій.
Теми	Подано нижче у таблиці Схема курсу «Вступ до машинного навчання».
Підсумковий контроль, форма	Залік
Пререквізити	Для вивчення курсу студенти потребують базових знань з

	<ul style="list-style-type: none"> - Основ програмування - Програмного забезпечення 																																										
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Презентації, лекції (лекція-розповідь, лекція-бесіда), модульний контроль. Домашні та індивідуальні завдання																																										
Необхідне обладнання	Комп'ютер із програмним забезпеченням Visual Studio 2017/2019, Visual Studio Code.																																										
Критерії оцінювання (окремо для кожного виду навчальної діяльності)	<p>Оцінювання проводиться за 100-бальною шкалою.</p> <table border="1"> <thead> <tr> <th colspan="2">Оцінка за шкалою ECTS</th> <th>Оцінка в балах</th> <th colspan="3">Оцінка за національною шкалою</th> </tr> <tr> <th colspan="2"></th> <th></th> <th colspan="2">Екзамен, диференційований залік</th> <th>залік</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>Відмінно</td> <td>100 - 90</td> <td>Відмінно</td> <td>5</td> <td rowspan="5">зараховано</td> </tr> <tr> <td>B</td> <td>Дуже добре</td> <td>81- 89</td> <td rowspan="2">Добре</td> <td>4</td> </tr> <tr> <td>C</td> <td>Добре</td> <td>71 -80</td> <td>3</td> </tr> <tr> <td>D</td> <td>Задовільно</td> <td>61 - 70</td> <td rowspan="2">Задовільно</td> <td>2</td> </tr> <tr> <td>E</td> <td>Достатньо</td> <td>51- 60</td> <td>1</td> </tr> <tr> <td>FX (F)</td> <td>Незадовільно</td> <td>0 - 50</td> <td>Незадовільно</td> <td>2</td> <td>не зараховано</td> </tr> </tbody> </table> <p>Впродовж семестру студент може отримати 100 балів. З них:</p> <ul style="list-style-type: none"> - домашні завдання: 25% семестрової оцінки; максимальна кількість балів 25 (5 завдання по 5 балів); - індивідуальні завдання: 30% семестрової оцінки; максимальна кількість балів 30 (2 завдання по 15 балів); завдання оформлено на основі індивідуальних наборів даних з платформи Kaggle; для кожного завдання встановлено терміни здачі. Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку (кожного лабораторного заняття після терміну здачі на 2 бали менше); - контрольні заміри (модулі): 45% семестрової оцінки; максимальна кількість балів 45 (3 тестові модулі в середовищі MS Teams по 15 балів); <p>Підсумкова максимальна кількість балів 100.</p> <p>Критерії оцінювання домашніх завдань: 5 балів – студент повністю виконав умови завдання; алгоритм реалізовано правильно; продемонстровано приклади використання; оформлено належним чином за допомогою Jupyter Notebook з коментарями та візуалізаціями; 3-4 бали – завдання реалізовано повністю, але з незначними помилками; оформлено не в повній відповідності до вимог: частково відсутні візуалізації або коментарі; 1-2 бали – завдання реалізовано частково або із значними помилками; відсутні приклади використання; оформлено неналежним чином; 0 балів - студент не виконав завдання.</p> <p>Критерії оцінювання індивідуальних завдань: 15 балів – студент повністю виконав умови завдання; проведено</p>	Оцінка за шкалою ECTS		Оцінка в балах	Оцінка за національною шкалою						Екзамен, диференційований залік		залік	A	Відмінно	100 - 90	Відмінно	5	зараховано	B	Дуже добре	81- 89	Добре	4	C	Добре	71 -80	3	D	Задовільно	61 - 70	Задовільно	2	E	Достатньо	51- 60	1	FX (F)	Незадовільно	0 - 50	Незадовільно	2	не зараховано
Оцінка за шкалою ECTS		Оцінка в балах	Оцінка за національною шкалою																																								
			Екзамен, диференційований залік		залік																																						
A	Відмінно	100 - 90	Відмінно	5	зараховано																																						
B	Дуже добре	81- 89	Добре	4																																							
C	Добре	71 -80		3																																							
D	Задовільно	61 - 70	Задовільно	2																																							
E	Достатньо	51- 60		1																																							
FX (F)	Незадовільно	0 - 50	Незадовільно	2	не зараховано																																						

	<p>попередній аналіз даних та побудовано візуалізації; здійснено вибір та порівняння архітектур для розв'язування завдання; підбрано оптимальні гіперпараметри для алгоритму; проведено аналіз та візуалізацію результатів; остаточно модель показує високий результат;</p> <p>10-14 балів – студент повністю виконав умови завдання, але з незначними помилками або зауваженнями щодо оформлення роботи;</p> <p>5-9 балів – студент виконав завдання частково, але при цьому зміг побудувати кінцеву модель, яка показує хороший результат; відсутній один з обов'язкових пунктів роботи: дослідження та візуалізації даних, підбору гіперпараметрів алгоритму або аналізу результатів</p> <p>1-4 бали – студент виконав завдання частково з грубими помилками; присутні помилки виконання програми; відсутня кінцева модель для задачі;</p> <p>0 балів – студент не виконав завдання.</p> <p>Критерії оцінювання контрольних замірів (модуль):</p> <p>1 бал – відповідь на запитання правильна;</p> <p>0 балів – відповідь на завдання неправильна.</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>Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції та лабораторні зайняття курсу. Студенти повинні інформувати викладача про неможливість відвідати заняття. У будь-якому випадку студенти зобов'язані дотримуватися термінів визначених для виконання домашніх 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 на лабораторних заняттях (здача домашніх та індивідуальних завдань) та під час контрольних замірів. При цьому обов'язково враховуються присутність на заняттях та активність студента під час лабораторного 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>Питання до контрольних замірів.</p>	<p>Лінійна регресія:</p> <ul style="list-style-type: none"> - функція вартості - метод найшвидшого спуску - нормалізація характеристик - швидкість навчання - регуляризація - методи оцінки моделі <p>Задачі класифікації:</p> <ul style="list-style-type: none"> - логістична регресія; представлення гіпотези; функція вартості;

	<p>градієнтний спуск</p> <ul style="list-style-type: none"> - точність моделей МН - бассівський наївний класифікатор - нейронні мережі; модель нейрона; приховані шари; прямий та зворотній хід нейронної мережі - регуляризація та оптимізація нейронних мереж - задача класифікації тексту <p>Неконтрольоване машинне навчання:</p> <ul style="list-style-type: none"> - задача кластеризації; алгоритм k-середніх - задача зменшення розмірності даних; основні алгоритми - задача виявлення аномалій; основні алгоритми
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

Схема курсу «Вступ до машинного навчання»

Тижень	Тема, план, короткі тези	Форма діяльності (заняття)	Література, Ресурси в інтернеті	Завдання, год.	Термін виконання
1	Тема 1. Предмет дисципліни (<i>Означення. Типи машинного навчання. Історія. Приклади</i>)	Лекція (2 год.)	[1], [3-4]	Опрацювання лекційного матеріалу (3 год.)	1 тиждень
	Тема 1. Елементи лінійної алгебри (<i>Програмування матричних операцій. Використання питру та інших бібліотек для аналізу даних</i>)	Лабораторне (2 год.)	[2], [5]		1 тиждень
2	Тема 2. Лінійна регресія з однією змінною (<i>Модель. Функція вартості. Метод градієнтного спуску</i>)	Лекція (2 год.)	[1], [3-4]	Опрацювання лекційного матеріалу (2 год.)	1 тиждень
	Тема 2. Програмування алгоритму лінійної регресії для однієї змінної Домашнє завдання №1. Реалізувати векторизований алгоритм лінійної регресії та розв'язати поставлені задачі	Лабораторне (2 год.)	[5]	Виконання домашнього завдання №1 (5 год.)	2 тижні
3	Тема 3. Лінійна регресія з багатьма змінними (<i>Побудова моделі. Нормалізація рис. Швидкість навчання</i>)	Лекція (2 год.)	[1], [3-4]	Опрацювання лекційного матеріалу (2 год.)	1 тиждень
	Тема 3. Модифікація алгоритму лінійної регресії	Лабораторне (2 год.)	[5]	Підготовка доповіді за	5 тижнів

	для випадку багатьох змінних. Індивідуальне завдання №1. Розв'язування задачі регресії з платформи Kaggle			результатами (7 год.)	
4	Тема 4. Задачі класифікації (<i>Логістична регресія. Представлення гіпотези. Функція вартості. Градієнтний спуск</i>)	Лекція (2 год.)	[1], [3-4]	Опрацювання лекційного матеріалу (2 год.)	1 тиждень
	Тема 4. Розв'язування модельних задач класифікації. Здача домашнього завдання №1. Домашнє завдання №2. Реалізувати векторизований алгоритм логістичної регресії та розв'язати поставлені задачі	Лабораторне (2 год.)	[5]	Виконання завдання №2 (5 год.)	2 тижні
5	Тема 5. Точність моделей (<i>Проблема перевизначення моделі. Регуляризація лінійної та логістичної регресії</i>)	Лекція (2 год.)	[1], [3-4]	Опрацювання лекційного матеріалу (3 год.)	1 тиждень
	Тема 5. Використання поліноміальних характеристик для розв'язування задачі класифікації	Лабораторне (2 год.)	[5]		1 тиждень
6	Тема 6. Імовірнісний підхід (<i>Баєсівський наївний класифікатор</i>) Здача домашнього завдання №2.	Лекція (2 год.)	[1], [3-4]	Опрацювання лекційного матеріалу (2 год.)	1 тиждень
	Тема 6. Програмування Баєсівського наївного класифікатора. Контрольний замір №1. Індивідуальне завдання №2. Розв'язування задачі класифікації з платформи Kaggle	Лабораторне (2 год.)	[5]	Підготовка доповіді за результатами (8 год.)	8 тижнів
7	Тема 7. Нейронні мережі (<i>Зв'язок із логістичною регресією. Представлення</i>)	Лекція (2 год.)	[1], [6-9]	Опрацювання лекційного матеріалу	1 тиждень

	<i>моделі. Приховані шари. Робота алгоритму)</i>			(2 год.)	
	Тема 7. Реалізація алгоритму прямого та зворотного ходу для одношарової нейронної мережі. Домашнє завдання №3. Реалізувати алгоритм нейронної мережі для розв'язування задач багатокласової класифікації	Лабораторне (2 год.)	[5]	Виконання завдання №3 (8 год)	4 тижні
8	Тема 8. Нейронні мережі (<i>Багатокласова класифікація. Softmax. Ідея реалізації алгоритму</i>)	Лекція (2 год.)	[1], [6-9]	Опрацювання лекційного матеріалу (3 год.)	1 тиждень
	Тема 8. Використання класу MLPClassifier з бібліотеки scikit-learn для розв'язування модельних задач класифікації Здача індивідуального завдання №1.	Лабораторне (2 год.)	[5]		1 тиждень
9	Тема 9. Нейронні мережі (<i>Регуляризація та оптимізація роботи нейронних мереж: стохастичний градієнтний спуск, Метод моментів, метод Адама</i>)	Лекція (2 год.)	[1], [6-9]	Опрацювання лекційного матеріалу (2 год.)	1 тиждень
	Тема 9. Приклад роботи нейронної мережі на даних з бази Kaggle. Домашнє завдання №4. Реалізувати алгоритм оптимізації нейронної мережі	Лабораторне (2 год.)	[5]	Виконання завдання №4 (6 год)	2 тижні
10	Тема 10. Задача класифікації тексту (<i>Обробка тексту: стоп-слова, лематизація, нормалізація. Tf-Idf. Основні алгоритми</i>)	Лекція (2 год.)	[1], [6-9], [11]	Опрацювання лекційного матеріалу (3 год.)	1 тиждень
	Тема 10. Задачі класифікації з текстовими характеристиками. Використання методів з модуля scikit-learn.	Лабораторне (2 год.)	[5], [11]		1 тиждень
11	Тема 11. Дизайн систем контролюваного	Лекція (2 год.)	[1], [3-4], [6-9]	Опрацювання лекційного	1 тиждень

	машинного навчання (Аналіз даних. Вибір моделі та метрики продуктивності. Аналіз та покращення результатів роботи алгоритму)			матеріалу (3 год.)	
	Тема 11. Системи підбору гіперпараметрів моделей (Grid search, randomized search) Здача домашніх завдань №3-4.	Лабораторне (2 год.)	[5]		1 тиждень
12	Тема 12. Сучасні задачі машинного навчання (Згорткові та рекурентні мережі. Ідеї алгоритмів. Застосування)	Лекція (2 год.)	[6-11]	Опрацювання лекційного матеріалу (3 год.)	1 тиждень
	Тема 12. Класифікація зображень за допомогою згорткових нейронних мереж Контрольний замір №2.	Лабораторне (2 год.)	[6-10]		1 тиждень
13	Тема 13. Неконтрольоване машинне навчання (Задача кластеризації. Постановка задачі. Алгоритм K-середніх. Обрання кількості кластерів)	Лекція (2 год.)	[1], [3-4]	Опрацювання лекційного матеріалу (2 год.)	1 тиждень
	Тема 13. Використання бібліотеки scikit-learn для розв'язування задач кластеризації. Домашнє завдання №5. Реалізувати алгоритм K середніх	Лабораторне (2 год.)	[5]	Виконання завдання №5 (6 год.)	3 тижні
14	Тема 14. Задача зменшення розмірності даних (Основні алгоритми. Використання)	Лекція (2 год.)	[1], [3-4]	Опрацювання лекційного матеріалу (3 год.)	1 тиждень
	Тема 14. Робота з алгоритмами зменшення розмірності даних на демонстраційних датасетах за допомогою scikit-learn. Здача індивідуального завдання №2.	Лабораторне (2 год.)	[5]		1 тиждень
15	Тема 15. Виявлення аномалій (Постановка задачі. Гаусівський	Лекція (2 год.)	[1], [3-4]	Опрацювання лекційного матеріалу	1 тиждень

	<i>розподіл. Алгоритми. Обрання характеристик)</i>			(3 год.)	
	Тема 15. Робота з алгоритмами виявлення аномалій на демонстраційних датасетах за допомогою scikit-learn.	Лабораторне (2 год.)	[5]		1 тиждень
16	Тема 16. Системи автоматичного тренування моделей машинного навчання (<i>Використання хмарних технологій</i>)	Лекція (2 год.)	[1], [3-4]	Опрацювання лекційного матеріалу (3 год.)	1 тиждень
	Контрольний замір №3. Здача домашнього завдання №5.	Лабораторне (2 год.)			