



ОСНОВИ МАШИННОГО НАВЧАННЯ

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Перший (бакалаврський)</i>
Галузь знань	<i>11 Математика і статистика</i>
Спеціальність	<i>113 Прикладна математика</i>
Освітня програма	<i>Наука про дані та математичне моделювання</i>
Статус дисципліни	<i>Нормативна</i>
Форма навчання	<i>Очна (денна)</i>
Рік підготовки, семестр	<i>3 курс, весняний семестр</i>
Обсяг дисципліни	<i>105 год. / 3,5 кредити, лекції 18 год., лабораторні 36 год.</i>
Семестровий контроль/ контрольні заходи	<i>Залік/контрольна робота у формі колоквиума, лабораторні роботи</i>
Розклад занять	<i>Лекція – 1 год. на тиждень, лабораторні роботи – 2 год. на тиждень</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: проф., д.т.н. Чертов Олег Романович, chertov@i.ua Лабораторні: Жук Іван Сергійович, ivan.sergeyevich.zhuk@gmail.com</i>
Розміщення курсу	

Програма навчальної дисципліни

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

Метою навчальної дисципліни є формування у студентів сучасного наукового світогляду в області методів машинного навчання, ознайомлення із рядом моделей машинного навчання та роз'яснення аспектів їх застосування.

Предметом навчальної дисципліни є методи контрольованого та неконтрольованого навчання, перехресної перевірки та відбору важливих характеристик.

Передбачається отримання здобувачами таких компетенцій та результатів навчання (у застосуванні до сфери машинного навчання):

- здатність учитися і оволодівати сучасними знаннями;
- знання та розуміння предметної області та розуміння професійної діяльності;
- здатність використовувати й адаптувати математичні теорії, методи та прийоми для доведення математичних тверджень і теорем;
- здатність виконувати завдання, сформульовані у математичній формі;
- здатність сформулювати математичну постановку задачі, спираючись на постановку мовою предметної галузі, та обирати метод її розв'язання, що забезпечує потрібні точність і надійність результату;
- демонструвати знання й розуміння основних концепцій, принципів, теорій прикладної математики і використовувати їх на практиці;

- володіти основними положеннями та методами математичного, комплексного та функціонального аналізу, лінійної алгебри та теорії чисел, аналітичної геометрії, теорії диференціальних рівнянь, зокрема рівнянь у частинних похідних, теорії ймовірностей, математичної статистики та випадкових процесів, чисельними методами;
- формалізувати задачі, сформульовані мовою певної предметної галузі; формулювати їх математичну постановку та обирати раціональний метод вирішення; розв'язувати отримані задачі аналітичними та чисельними методами, оцінювати точність та достовірність отриманих результатів;
- уміти розробляти та використовувати на практиці алгоритми, пов'язані з апроксимацією функціональних залежностей, чисельним диференціюванням та інтегруванням, розв'язанням систем алгебраїчних, диференціальних та інтегральних рівнянь, розв'язанням крайових задач, пошуком оптимальних рішень;
- вміти формулювати та розв'язувати задачі з динаміки матеріальної точки;
- вміти формулювати та розв'язувати задачі аналізу даних і базові задачі машинного навчання.

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

Необхідно опанування університетських курсів алгебри, математичного аналізу, дискретної математики, теорії ймовірностей, математичної статистики, методів оптимізації, програмування мовою Python.

3. Зміст навчальної дисципліни

Розділ 1. Контрольоване навчання

- Тема 1.1. Вступ
- Тема 1.2. Класифікаційні метрики
- Тема 1.3. Дерева прийняття рішень
- Тема 1.4. Метричні методи класифікації та регресії

Розділ 2. Неконтрольоване навчання. Складні моделі

- Тема 2.1. Кластеризація
- Тема 2.2. Інженерія інформативних характеристик
- Тема 2.3. Комбінація моделей

4. Навчальні матеріали та ресурси

Базова література

1. Кононова К. Ю. Машинне навчання: методи та моделі. – Харків: ХНУ імені В. Н. Каразіна, 2020. – 301 с. (<https://github.com/katerynakononova/ML/blob/master/ML.pdf>)
2. Інтелектуальний аналіз даних та машинне навчання. Ч. 1 : Базові методи та засоби аналізу даних / Я. В. Іванчук та ін. / Вінниц. нац. техн. ун-т. - Вінниця : ВНТУ, 2021. - 68 с. (http://pdf.lib.vntu.edu.ua/books/2022/lvanchuk_P1_2021_69.pdf).

Допоміжна література

1. Burkov, A. (2019) The Hundred-Page Machine Learning Book (<https://themlbook.com/>).
2. James, G.M., Witten, D.M., Hastie, T.J., & Tibshirani, R. (2021). An Introduction to Statistical Learning. Springer Texts in Statistics (https://hastie.su.domains/ISLR2/ISLRv2_website.pdf).

Інформаційні ресурси

1. Електронний кампус КПІ ім. Ігоря Сікорського. Матеріали з дисципліни «Основи машинного навчання». – Режим доступу : <http://login.kpi.ua>
2. Дистанційний курс “Машинне навчання” - Електронний ресурс — Режим доступу https://courses.prometheus.org.ua/courses/IRF/ML101/2016_T3/about

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

5.1. Лекції

Лекція 1 Вступ до машинного навчання та науки про дані

Визначення науки про дані (через діаграми Венна). Визначення машинного навчання (за Томом Мітчелом). Типи навчання та класи задач машинного навчання. Базова схема машинного навчання.

Лекції 2-3 Класифікаційні метрики

Класифікаційні метрики: матриця помилок (Confusion matrix) і помилки першого та другого роду, доля правильних відповідей (Accuracy), точність (Precision) і повнота (Recall), F-міри (F-Scores), коефіцієнт кореляції Метьюса (Matthews Correlation Coefficient, MCC), Precision-Recall крива (curve), ROC («reciever operating characteristic»)-крива, Area under curve (AUC). Бінарний та багатокласовий випадки.

Лекції 4 Древа прийняття рішень

Використання дерев прийняття рішень для класифікації. Експоненційний розмір простору гіпотез. Класифікація неперервних даних за допомогою дерев. Жадібний алгоритм побудови дерева. Проблеми при наявності атрибутів з багатьма значеннями. Побудова по дереву правил прийняття рішень.

Алгоритм ID3.

Порівняння критеріїв розщеплення:

- *Information Gain (приріст інформації)*
- *Gini impurity (неоднорідність Джині)*
- *(Information) Gain ratio (нормалізований приріст інформації)*
- *Classification error (помилка класифікації).*

Переваги та недоліки дерев прийняття рішень.

Обрізання дерев (pre-pruning, post-pruning). Для чого воно застосовується?

Лекції 5 Метричні методи класифікації

Відмінність задач регресії та класифікації. Метрики, що базуються на відстані та подібності (distance-based vs similarity-based). Жадібні / ледачі алгоритми. Гіпотези компактності та неперервності. Алгоритм k найближчих сусідів. Яке значення k краще? Підбір вагів. Діаграма Вороного. Чутливість до шумів. Нормалізація характеристик. Приклади метрик (метрика Мінковського, косинусна міра, відстань Жаккарда, відстань Хеммінга).

Лекції 6 Кластеризація

Кластеризація як приклад навчання без вчителя. Типи кластерних алгоритмів.

Характеристики кластерів. Критерії зупинки кластеризації. Метод K-середніх. Як визначити кількість кластерів? Ієрархічна агломеративна кластеризація.

Лекція 7 Інженерія інформативних характеристик

Відбір інформативних характеристик (feature selection): виявлення неінформативних ознак за допомогою фільтрів; методи обгортки; методи, вбудовані в модель навчання.

Лекції 8 Комбінації моделей

Загальна схема ансамблю моделей. За рахунок чого є перевага ансамблю над окремим класифікатором? *Bagging*. Приклади агрегуючих функцій. Просте і зважене голосування. Випадковий ліс. *Boosting* (підсилення).

Лекція 9 Типовий підхід до розв'язання задач машинного навчання

5.2. Лабораторні роботи (орієнтовна тематика)

- 1 Аналіз даних із використанням Python.
- 2 Застосування метрик класифікації.
- 3 Дерева прийняття рішень.
- 4 Метричні методи класифікації та регресії. Алгоритм k найближчих сусідів.
- 5 Кластеризація.

6. Самостійна робота студента

- | | | |
|---|--|---------|
| 1 | Підготовка до захисту лабораторних робіт
<i>7 годин на підготовку до кожної лабораторної роботи</i> | 35 год. |
| 2 | Підготовка до модульної контрольної роботи | 6 год. |
| 3 | Підготовка до заліку | 10 год. |

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

- Відвідування лекцій та лабораторних занять є дуже бажаним.
- На лекційних заняттях студенти повинні слухати уважно, не створювати шуму. Дозволяється (і рекомендується!) задавати «гарні» питання для з'ясування незрозумілих моментів та відповідати на запитання викладача. Мобільні телефони, планшети, смарт-годинники та інші гаджети мають бути вимкнені або встановлені на беззвучний режим, користуватися ними не дозволяється, якщо вони не застосовуються для організації навчання.
- Лабораторні роботи виконуються за допомогою мови Python, можливість використання сторонніх бібліотек визначається для кожної лабораторної роботи окремо. Захист лабораторних робіт відбувається в усній формі, очно або в режимі конференції (при дистанційній формі навчання). На захисті студент має розповісти хід виконання роботи, сформулювати висновки, відповідати на питання викладача. Роботи потрібно здавати та захищати вчасно, за запізнення оцінка знижується.
- Модульна контрольна робота виконується під час одного із лекційних або лабораторних занять у формі колоквіума.
- Під час вивчення дисципліни студенти повинні дотримуватися правил академічної доброчесності, що передбачає зокрема неприпустимість плагіату, списування та інших способів видавання чужого доробку за свій. За недотримання цих правил передбачається покарання, що включає в себе вилучення порушника/порушників з відповідного контрольного заходу та виставлення за нього оцінки «0». Якщо при перевірці лабораторних робіт або модульної контрольної роботи виявляються ознаки списування, оцінка «0» виставляється і тому, хто списав, і тому, хто дав списати.

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

Рейтинг студента з дисципліни складається з балів, що він отримує:

- 1) за виконання, здачу та захист лабораторних робіт;
- 2) за відповіді на колоквіумі (як виконання модульної контрольної роботи, МКР);

- 3) як заохочувальні за виконання творчих чи оригінальних робіт з дисципліни, за активну роботу на лекціях.

Система рейтингових (вагових) балів та критерії оцінювання

1. Виконання, здача та захист лабораторних робіт

За семестр студент повинен підготувати, здати і захистити викладачу 5 лабораторних робіт. При оцінюванні лабораторних робіт, як і при розв'язанні реальних практичних задач для конкретного Замовника, враховується працездатність програми, своєчасність її надання та вміння пояснити роботу програми й інтерпретувати отримані результати.

Окремо перевіряється академічна доброчесність – студент повинен довести викладачу, що він розуміє, що саме і як саме виконується в програмі та може в ній пояснити особливості коду і міркування, що лягли в основу прийнятих рішень при програмній реалізації.

За виконання та здачу однієї лабораторної роботи студент може отримати бали за кожен із наступних трьох складових:

- 1) (*своєчасність*): якщо лабораторна робота надіслана викладачу не пізніше обумовленого терміну і при цьому безпомилково проходить базове тестування (запускається без помилок та дає коректний результат хоча б для одного стандартного тестового прикладу), то за це нараховується 4 бали;
- 2) (*працездатність*):
 - якщо програма працює безпомилково, то за це нараховується 3 бали;
 - якщо програма має несуттєві недоліки, інколи дає неправильний результат, то за це нараховується від 1 до 2 балів;
 - якщо програма у більшості випадків працює некоректно, то за це нараховується 0 балів;
- 3) (*презентабельність*):
 - якщо студент правильно чи майже правильно відповідає на питання стосовно функціональності програми та особливостей її реалізації, то за це нараховується від 6 до 7 балів;
 - якщо студент частково плутається при відповіді на питання стосовно функціональності програми та особливостей її реалізації, то за це нараховується від 3 до 5 балів;
 - якщо студент частково плутається при відповіді на питання стосовно функціональності програми та не може обґрунтувати вибір інструментальних засобів для її реалізації, то за це нараховується від 1 до 2 балів;
 - якщо студент не може пояснити основну функціональність програми та особливостей її реалізації, тобто очевидно має місце порушення академічної доброчесності, то всі бали за відповідну лабораторну роботу обнуляються і, загалом, за неї студент отримує 0 балів;
 - щоб зрозуміти, наскільки студент розібрався з теоретичним матеріалом, який відповідає лабораторній роботі, викладач може задати студенту декілька питань по відповідній теорії, і в залежності від правильності відповідей на них збільшує чи зменшує оцінку за «презентабельність».

Таким чином, за одну лабораторну роботу один студент максимально може отримати:

$$4 \text{ бали} + 3 \text{ бали} + 7 \text{ балів} = \mathbf{14 \text{ балів.}}$$

Загалом, за цикл лабораторних робіт студент максимально може отримати:

$$14 \text{ балів} \times 5 = \mathbf{70 \text{ балів (} R_{Lab}\text{)}}.$$

Викладач має право заохотити студента певною кількістю балів (максимально – 5 балів за семестр) за надання цікавої і оригінальної алгоритмічної чи програмної реалізації завдання лабораторної роботи.

2. Бали за відповіді на колоквіумі

Колоквіум проводиться як усна співбесіда, на якій кожен студент повинен відповісти на три основні запитання від викладача. Для уточнення відповіді студента чи перевірки – наскільки глибоко він розуміє матеріал, викладач може задавати запитання, додаткові до основного.

Відповідь на кожне запитання оцінюється максимум у 10 балів.

Критерії оцінювання:

- 10 балів — відповідь повна і правильна;
- 6-9 балів — відповідь правильна, але неповна або містить незначні неточності;
- 1-5 бали — відповідь неповна і неточна;
- 0 балів — відповідь неправильна або відсутня.

Максимальна кількість балів за колоквіум:

10 балів × 3 запитання = **30 балів**.

3. Заохочувальні бали за:

- виступ на лекції (за попередньою домовленістю з лектором) із доповіддю по матеріалам навчального курсу (або близьким до них) чи з доповіддю про особисто отримані практичні чи наукові результати з машинного навчання надається до 5-ти заохочувальних балів;
- активність на лекціях, тобто за
 - відповіді на запитання лектора до загальної аудиторії,
 - знаходження описок/помилки у лекціях,
 - задавання «правильних запитань», тобто запитань, які свідчать про вдумливу роботу студента з навчальним матеріаломнараховується загалом до 5-ти заохочувальних балів.

4. Визнання результатів навчання, отриманих у неформальній освіті, тощо

За проходження певного елемента неформальної освіти (наприклад, онлайн-курсів Coursera або аналогічних), за перемогу або участь в тематичних хакатонах або інших змаганнях, студенту можуть бути нараховані додаткові бали, зараховано виконання лабораторних робіт за всю дисципліну з максимально можливою оцінкою, або, взагалі, виставлено 100 балів.

Конкретний розмір заохочення студента визначається викладачем, що читає лекції, виходячи із повноти, важливості та результатів проходження цим студентом відповідних елементів неформальної освіти. Але в будь-якому випадку визнання результатів навчання, отриманих у неформальній освіті, може бути зараховано лише у випадку попереднього узгодження з викладачем.

6. Розрахунок шкали (R) рейтингу

Сума вагових балів контрольних заходів протягом семестру складає:

$$R_C = R_{Lab} + R_{CW} = 70 \text{ балів} + 30 \text{ балів} = 100 \text{ балів.}$$

Неодмінною умовою допуску до заліку є набрання студентом 30 балів за семестр. Інакше студент у першу відомість отримує оцінку «не допущено», а потім повинен виконати ці умови (набирати бали для допуску). Виконавши умови допуску до заліку, студент під час першого та/або другого перескладання пише залікову контрольну роботу.

Якщо наприкінці семестру студент набрав не менше 60 рейтингових балів, а також виконав умови допуску до заліку, він отримує залік «автоматом» відповідно до таблиці.

Студент може спробувати підвищити свою оцінку шляхом написання залікової контрольної роботи, при цьому його бали, отримані за семестр, анулюються.

У випадку, якщо сума рейтингових балів менша від 60, але виконано умови допуску до семестрового контролю, студент виконує залікову контрольну роботу, при цьому його рейтинг анулюють, після чого бали нараховують за результатами виконання залікової контрольної роботи.

Залікова контрольна робота проводиться як колоквиум з 5 письмовими та усними завданнями. Відповідь на кожне завдання оцінюють у 20 балів.

Критерії оцінювання відповідей на кожне завдання:

20 балів — правильна та змістовна відповідь;

15–19 балів — відповідь правильна, змістовна, але має незначні вади;

10–14 балів — відповідь містить незначні помилки чи є неповною;

1–9 балів — відповідь містить суттєві помилки та є неповною;

0 балів — немає відповіді.

Для отримання студентом відповідних оцінок його рейтингова оцінка R_D переводиться згідно з такою таблицею:

Рейтингові бали, R_D	Оцінка
95–100	відмінно
85–94	дуже добре
75–84	добре
65–74	задовільно
60–64	достатньо
Сума балів < 60	незадовільно
Рейтинг за семестр < 30	не допущено

Робочу програму навчальної дисципліни (силабус):

Складено завідувачем кафедри ПМА, д.т.н., проф. ЧЕРТОВИМ О. Р.

Ухвалено кафедрою ПМА (протокол № 18 від 10.06.2024).

Погоджено Методичною комісією факультету (протокол № 12 від 21.06.2024).

Уточнення дат здачі етапів міні-проєкту (02.09.2024).