

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Машинне навчання



Ступінь освіти	Магістр
Освітня програма	Системний аналіз
Тривалість викладання	II-III чверті
Заняття:	II семестр
Лекції	2 год/тиждень
Практичні	1 год/тиждень
Мова викладання	українська, англійська

Кафедра, що викладає: системного аналізу та управління

Сторінка курсу в СДО НТУ «ДП»:

<https://do.nmu.org.ua/course/view.php?id=6019>

Консультації: за окремим розкладом, погодженим зі здобувачами вищої освіти

Онлайн-консультації: електронна пошта

Інформація про викладачів



Владико Олександр Борисович
(лекційні заняття, практичні заняття)

доцент, канд. техн. наук, доцент

Персональна сторінка

<http://prn.nmu.org.ua/index.php/sp/dot/3-vladiko-o-b>

E-mail

vladyko.o.b@nmu.one

1. Анотація до курсу

Особливістю методів машинного навчання є не прямий розв'язок задачі, а навчання на множині подібних прикладів, що дозволяє використовувати ці методи для обробки великих обсягів даних та виявляти в них нові, нетривіальні, корисні та доступні для інтерпретації знання. Окрім методів штучного інтелекту, під час розробки моделей машинного навчання в якості допоміжних використовуються засоби математичної статистики, чисельних методів, методів оптимізації, теорії ймовірностей, теорії графів, різні техніки роботи з даними в цифровій формі. Машинне навчання застосовують в ряді обчислювальних задач, в яких розробка та програмування явних алгоритмів з доброю продуктивністю є складною або нездійсненною задачею. Для практичного засвоєння навчальних матеріалів ряд тем дисципліни поглиблено вивчається на лабораторних заняттях.

2. Мета та завдання навчальної дисципліни

Мета дисципліни – сформувати у магістрів навички застосування та використання сучасних технологій обробки даних для вирішення задач класифікації, регресійного аналізу, прогнозування та ухвалення рішень.

Завдання курсу:

- теоретичних основ базових технологій машинного навчання;
- методів, алгоритмів роботи, засобів реалізації та технологій налагодження систем обробки даних з використанням технологій обчислювального інтелекту;
- сучасних програмних засобів реалізації технологій машинного навчання.

3. Результати навчання

Знають теоретичні основи базових технологій машинного навчання; методів, алгоритмів роботи, засобів реалізації та технологій налагодження систем обробки даних з використанням технологій обчислювального інтелекту; сучасних програмних засобів реалізації технологій машинного навчання.

Вміють обґрунтовано вибирати конкретні технології та алгоритми машинного навчання при розв'язанні відповідних практичних задач; здійснювати підготовку та первинну обробку даних для побудови моделей систем методами машинного навчання; вирішувати задачі автоматизації підтримки рішень, розпізнавання образів, діагностики, класифікації, оптимізації та аналізу даних методами машинного навчання; використовувати сучасні програмні засоби для реалізації технологій машинного навчання.

4. Структура курсу

Лекційні заняття

1. Машинне навчання. Вступ.

Постановка задачі машинного навчання. Лінійні моделі. Поняття перенавчання. Способи оцінювання ступеня перенавчання моделей.

2. Задачі бінарної класифікації.

Рівняння гіперплощини у задачах бінарної класифікації. Розв'язання простого завдання бінарної класифікації. Функції втрат у задачах лінійної бінарної класифікації.

3. Формула Байєса

Формула Байєса під час вирішення конкретних завдань. Байєсовський висновок. Наївна байєсівська класифікація. Гаусівський байєсівський класифікатор. Лінійний дискримінант Фішера.

4. Методи SVM, PCA

Введення метод опорних векторів (SVM) Реалізація методу опорних векторів (SVM) Метод опорних векторів (SVM) із нелінійними ядрами. Імовірна оцінка якості моделей

5. Методи головних компонентів

Показники precision та recall. F-міра. Метрики якості ранжирування. ROC-крива. Метод головних компонентів (Principal Component Analysis). Скорочення розмірності простору ознак за допомогою PCA. Сингулярне розкладання та його зв'язок з PCA.

5. Методи класифікації

Багатокласова класифікація. Методи one-vs-all та all-vs-all. Метричні методи класифікації. Метод k найближчих сусідів. Методи парзенівського вікна та потенційних функцій. Метричні регресійні методи. Формула Надарая-Ватсона.

6. Алгоритми кластеризації

Завдання кластеризації. Постановка задачі. Алгоритм кластеризації Ллойда (K-середніх, K-means). Алгоритм кластеризації DBSCAN. Агломеративна ієрархічна кластеризація. Дендограма.

7. Вирішальні дерева

Логічні методи класифікації. Критерії якості для побудови вирішальних дерев. Побудова вирішальних дерев жадібним алгоритмом ID3. Усічення (pruning) дерева, обробка перепусток та категоріальних ознак. Вирішальні дерева у задачах регресії. Алгоритм CART. Випадкові дерева та випадковий ліс. Бутстреп та бегінг.

8. Бустінг

Введення у бустінг (boosting). Алгоритм AdaBoost під час класифікації. Алгоритм AdaBoost у задачах регресії. Градієнтний бустінг та стохастичний градієнтний бустінг. Нейронні сітки. Короткий вступ до теорії. Навчання нейронної мережі. Алгоритм back propagation.

Практичні заняття

ПР 1. Реалізація гаусівського байєсівського класифікатора

ПР 2. Апроксимація даних ядерним згладжуванням

ПР 3. Алгоритм кластеризації Ллойда (*K*-середніх)

ПР 4. Алгоритм класифікації AdaBoost на вирішальних деревах

5. Технічне обладнання та/або програмне забезпечення

№ роботи (шифр)	Назва роботи	Інструменти, обладнання та програмне забезпечення, що застосовуються при проведенні роботи
1	Завдання лінійної бінарної класифікації.	Персональний комп'ютер, Microsoft Excel, Microsoft Word, PTC Mathcad, Активований аккаунт університетської пошти (student.i.p.@nmu.one) на Ofіс 365
2	Реалізація гаусівського байєсівського класифікатора.	
3	Апроксимація даних ядерним згладжуванням	
4	Алгоритм кластеризації Ллойда (<i>K</i> -середніх).	
5	Алгоритм класифікації CART на вирішальних деревах.	
6	Алгоритм класифікації AdaBoost на вирішальних деревах.	

6. Система оцінювання та вимоги

6.1. Навчальні досягнення здобувачів вищої освіти за результатами вивчення курсу оцінюватимуться за шкалою, що наведена нижче:

Рейтингова шкала	Інституційна шкала
90 – 100	відмінно
75-89	добре
60-74	задовільно
0-59	незадовільно

6.2. Здобувач вищої освіти може отримати підсумкову оцінку з навчальної дисципліни на підставі поточного оцінювання знань за умови, якщо набрана кількість балів з поточного тестування та самостійної роботи складатиме не менше 60 балів.

Максимальне оцінювання:

Теоретична частина	Практична частина		Бонус	Разом
	При своєчасному складанні	При несвоєчасному складанні		
50	50	35	15	100

Практичні роботи приймаються у вигляді звіту та за контрольними запитаннями до кожної з них.

Теоретична частина оцінюється за результатами здачі контрольної роботи, яка містить 4 питання та 4 задачі.

6.3. Критерії оцінювання елементів поточного та підсумкового контролю.

Кожна правильна відповідь з контрольних питань оцінюється у 15 балів (разом 40 балів). Контрольні питання проводиться з використанням технології Microsoft Forms Office 365. Написана на папері відповідь на контрольні питання скануються (фотографуються) та надсилається на електронну пошту викладача впродовж часу, відведеного на задачу теоретичної частини. Або, якщо дозволяє сервіс Microsoft Forms Office 365, набираються текстом у відповідні поля.

Задачі наводяться також у системі Microsoft Forms Office 365. Вирішена на папері задача сканується (фотографується) та надсилається на електронну пошту викладача впродовж часу, відведеного на задачу теоретичної частини. Несвоєчасно вислана відповідь враховується такою, що не здана.

Правильно вирішена **задача** оцінюється в 5 балів, причому:

- **5 балів** – відповідність еталону, з одиницями виміру;
- **4 бали** – відповідність еталону, без одиниць виміру або помилками в розрахунках;
- **3 бали** – незначні помилки у формулах, без одиниць виміру;
- **2 бали** – присутні суттєві помилки у рішенні;
- **1 бал** – наведені формули повністю не відповідають еталону;
- **0 балів** – рішення не наведене.

6.4. Критерії оцінювання практичної роботи

З кожної практичної роботи здобувач вищої освіти отримує 5 запитань з переліку контрольних запитань. Кількість вірних відповідей визначають кількість отриманих балів.

7. Політика курсу

7.1. Політика щодо академічної доброчесності

Академічна доброчесність здобувачів вищої освіти є важливою умовою для опанування результатами навчання за дисципліною і отримання задовільної оцінки з поточного та підсумкового контролів. Академічна доброчесність базується на засудженні практик списування (виконання письмових робіт із залученням зовнішніх джерел інформації, крім дозволених для використання), плагіату (відтворення опублікованих текстів інших авторів без зазначення авторства), фабрикації (вигадування даних чи фактів, що використовуються в освітньому процесі). Політика щодо академічної доброчесності регламентується положенням «Положення про систему запобігання та виявлення плагіату у Національному технічному університеті «Дніпровська політехніка».

У разі порушення здобувачем вищої освіти академічної доброчесності (списування, плагіат, фабрикація), робота оцінюється незадовільно та має бути виконана повторно. При цьому викладач залишає за собою право змінити тему завдання.

7.2. Комунікаційна політика

Здобувачі вищої освіти повинні мати активовану університетську пошту.

Усі письмові запитання до викладачів стосовно курсу мають надсилатися на університетську електронну пошту.

7.3. Політика щодо перескладання

Роботи, які здаються із порушенням термінів без поважних причин оцінюються на нижчу оцінку. Перескладання відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).

7.4. Політика щодо оскарження оцінювання

Якщо здобувач вищої освіти не згоден з оцінюванням його знань він може опротестувати виставлену викладачем оцінку у встановленому порядку.

7.5. Відвідування занять

Для здобувачів вищої освіти денної форми відвідування занять є обов'язковим. Поважними причинами для неявки на заняття є хвороба, участь в університетських заходах, академічна мобільність, які необхідно підтверджувати документами. Про відсутність на занятті та причини відсутності здобувач вищої освіти має повідомити викладача або особисто, або через старосту.

За об'єктивних причин (наприклад, міжнародна мобільність) навчання може відбуватись в онлайн формі за погодженням з керівником курсу.

8. Рекомендовані джерела інформації

8. Рекомендована література (базова)

1. Murphy, K. P. (2012). *Machine Learning: a Probabilistic Perspective*. MIT Press, Cambridge, MA, USA.
2. Farabet, C., LeCun, Y., Kavukcuoglu, K., Culurciello, E., Martini, B., Akselrod, P., and Talay, S. (2011). Large-scale FPGA-based convolutional networks. In R. Bekkerman, M. Bilenko, and J. Langford, editors, *Scaling up Machine Learning: Parallel and Distributed Approaches*. Cambridge University Press.
3. Кононова К. Ю. Машинне навчання: методи та моделі: підручник для бакалаврів, магістрів та докторів філософії спеціальності 051 «Економіка» / К. Ю. Кононова. – Харків: ХНУ імені В. Н. Каразіна, 2020. – 301 с.
4. Черняк А. И. Интеллектуальный анализ данных: учебник / А. И. Черняк, П. В. Захарченко; Киевский национальный университет імені Т. Шевченко. К.: Знання, 2014. 599 с.
5. Kevin P. Murphy [«Machine Learning: A Probabilistic Perspective»](#), 2012