

Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Кафедра прикладної математики

“ЗАТВЕРДЖУЮ”

Проректор
з науково-педагогічної роботи
Антон ПАНТЕЛЕЙМОНОВ



_____ 2020 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Прикладні задачі машинного навчання

(назва навчальної дисципліни)

рівень вищої освіти _____ магістр _____

галузь знань _____ 11 Математика та статистика _____

спеціальність _____ 113 Прикладна математика _____
(шифр і назва)

освітня програма _____ Прикладна математика _____
(шифр і назва)

спеціалізація _____
(шифр і назва)

вид дисципліни _____ за вибором _____
(обов'язкова / за вибором)

факультет _____ математики і інформатики _____

2020 / 2021 навчальний рік

Програму рекомендовано до затвердження Вченою радою факультету математики і інформатики

“31” серпня 2020 року, протокол № 8

РОЗРОБНИКИ ПРОГРАМИ: (вказати авторів, їхні наукові ступені, вчені звання та посади)
 Півень Олексій Леонідович, канд. фіз.-мат. наук, доцент кафедри прикладної математики.

Програму схвалено на засіданні кафедри Прикладної математики

Протокол від “31” серпня 2020 року №12

Завідувач кафедри Прикладної математики

ВК Валерій КОРОБОВ
 (підпис) (прізвище та ініціали)

Програму погоджено з гарантом освітньої освітньої-професійної програми (керівником проектної групи) Прикладна математика
 назва освітньої програми

Гарант освітньої-професійної програми
 (керівник проектної групи) Прикладна математика

Микола Микола ПАЦЕГОН
 (підпис) (прізвище та ініціали)

Програму погоджено науково-методичною комісією факультету математики і інформатики Харківського національного університету імені В.Н. Каразіна
 назва факультету, для здобувачів вищої освіти якого викладається навчальна дисципліна

Протокол від “31” серпня 2020 року, протокол №1

Голова науково-методичної комісії факультету математики і інформатики

Ольга Ольга АНОЦЕНКО
 (підпис) (прізвище та ініціали)

ВСТУП

Програма навчальної дисципліни “Прикладні задачі машинного навчання: моделювання та оптимізація” складена відповідно до освітньо-професійної програми підготовки

_____магістр_____ (назва рівня вищої освіти)

спеціальності (напряму) _____113 Прикладна математика_____

спеціалізації _____

1. Опис навчальної дисципліни

1.1. Мета викладання навчальної дисципліни

Метою викладання навчальної дисципліни “Прикладні задачі машинного навчання ” є навчання майбутніх спеціалістів сучасним методам побудови та аналізу різноманітних моделей машинного навчання, а також практичному застосуванню для вирішення деяких прикладних задач.

1.2. Основні завдання вивчення дисципліни.

Основними завданнями вивчення дисципліни “Прикладні задачі машинного навчання ” є

- ознайомити студентів із основними поняттями машинного навчання;
- вивчити основні сучасні методи, що використовуються при аналізі моделей;
- здобути практичні навички побудови та аналізу моделей.

1.3. Кількість кредитів - 6

1.4. Загальна кількість годин - 180

1.5. Характеристика навчальної дисципліни	
за вибором	
Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
1-й	-й
Семестр	
2-й	-й
Лекції	
32 год.	год.
Практичні, семінарські заняття	
32 год.	год.
Лабораторні заняття	
год.	год.
Самостійна робота	
116 год.	год.
у тому числі індивідуальні завдання	
год.	

1.6. Заплановані результати навчання

У результаті вивчення даного курсу студент повинен

знати:

- задачі та термінологію машинного навчання.
- основні моделі машинного навчання, що базуються на деревах прийняття рішень.
- множинну та логістичну регресію
- метод головних компонент
- байєсовські методи класифікації, зокрема методи дискримінантного аналізу
- метричні методи та метод опорних векторів класифікації об'єктів
- методи кластерного аналізу «класифікації без вчителя»

вміти:

- застосовувати для аналізу моделі, що базуються на деревах прийняття рішень.
- застосовувати для аналізу моделі, що базуються на множинній і логістичній регресії
- застосовувати метод головних компонент в задачах регресії для зниження розмірності та розв'язання проблеми мультиколінеарності
- застосовувати байєсовські методи, метричні методи, метод опорних векторів та методи кластерного аналізу для розв'язання задач класифікації об'єктів

2. Тематичний план навчальної дисципліни

Тема 1. Вступ до машинного навчання

Основні завдання курсу та поняття, пов'язані з машинним навчанням. Машинне навчання з вчителем та без. Ймовірнісна постановка задачі машинного навчання. Функціонали якості. Проблема перенавчання. Вступ у мову та середовище статистичної обробки даних R

Тема 2 Задачі регресії в машинному навчанні та метод головних компонент

Множинна регресія. Сингулярне розкладання матриць. Проблема мультиколінеарності в множинній регресії її розв'язання за допомоги методу головних компонент. Лінійні регресійні моделі зі змінною структурою. Фіктивні змінні.

Тема 3 Класифікація за допомоги логістичної регресії

Побудова логістичної регресії, оцінка параметрів за допомоги методу максимальної правдоподібності.

Тема 4. Моделі машинного навчання, що базуються на деревах прийняття рішень

Дерево класифікації, регресії. Використання ентропії Шеннона та індексу Gini.

Тема 5. Багатовимірний нормальний розподіл.

Означення багатовимірного нормального розподілу. Гіпотези про середнє, рівність середніх та коваріаційних матриць у багатовимірному нормальному розподілу.

Тема 6. Байєсовський метод класифікації

Ймовірна постановка задачі класифікації. Функціонал середнього ризику. Оптимальне байєсовське вирішальне правило, наївний байєсовський класифікатор. Дискримінантний аналіз Фішера: випадок нормального розподілу класів.

Тема 7. Метричні методи класифікації

Класифікація з вчителем. Міри схожості об'єктів. Метод найближчого сусіда та метод k-найближчих сусідів.

Тема 8. Кластерний аналіз

Класифікація без вчителя. Метод k-середніх та ієрархічна кластер-процедура.

Тема 9. Метод опорних векторів класифікації об'єктів.

Постановка задачі. Випадок лінійно розподіленої та лінійно нерозподіленої вибірки. Ядра та випрямляючі простори.

3. Структура навчальної дисципліни

Назви розділів і тем	Кількість годин											
	денна форма						заочна форма					
	усього	у тому числі					усього	у тому числі				
л		п	лаб.	інд.	с.р.	л		п	лаб.	інд.	с.р.	
1	2	3	4	5	6	7	8	9	10	11	12	13
Тема 1. Вступ до машинного навчання	20	4	4			12						
Тема 2. Задачі регресії в машинному навчанні та метод головних компонент	21	4	4			13						
Тема 3. Класифікація за допомоги логістичної регресії	17	2	2			13						
Тема 4. Моделі машинного навчання, що базуються на деревах прийняття рішень.	21	4	4			13						
Тема 5. Багатовимірний нормальний розподіл	21	4	4			13						
Тема 6. Байєсовський метод класифікації	21	4	4			13						
Тема 7. Метричні методи класифікації	21	4	4			13						
Тема 8. Кластерний аналіз	17	2	2			13						
Тема 9. Метод опорних векторів класифікації об'єктів	21	4	4			13						
Усього годин	180	32	32	0	0	116						

4. Теми семінарських (практичних, лабораторних) занять

№ з/п	Назва теми	Кількість годин
1	Вступ до машинного навчання. Статистичний аналіз в пакеті R	4
2	Задачі регресії в машинному навчанні та метод головних компонент	4
3	Класифікація за допомоги логістичної регресії	2
4	Моделі машинного навчання, що базуються на деревах прийняття рішень.	4
5	Багатовимірний нормальний розподіл	4
6	Байєсовський метод класифікації	4
7	Метричні методи класифікації	4
8	Кластерний аналіз	2
9	Метод опорних векторів класифікації об'єктів	2
10	<i>Контрольна робота</i>	2
	Разом	32

5. Завдання для самостійної роботи

№ з/п	Види, зміст самостійної роботи	Кількість годин
1	Знайомство з основними поняттями, пов'язаними з машинним навчанням	12
2	Знаходження головних компонент та побудова ліній регресії	13
3	Класифікація за допомоги логістичної регресії	13
4	Побудова дерев прийняття рішень та дерева регресії	13
5	Перевірка гіпотез про багатовимірний нормальний розподіл	13
6	Класифікація об'єктів за допомоги дискримінантного аналізу	13
7	Вивчення метричних методів класифікації	13
8	Розв'язання задач за допомогою кластерного аналізу	13
9	Класифікація об'єктів методом опорних векторів	13
	Разом	116

6. Індивідуальні завдання

7. Методи навчання

Лекційні та практичні заняття проводяться аудиторно. А у разі оголошення карантину заняття проводяться відповідно до Наказу ректора Харківського національного університету імені В.Н. Каразіна (аудиторно або дистанційно за допомогою платформ Google Meet або Zoom).

8. Методи контролю

- перевірка контрольної роботи;
- перевірка завдань для самостійної роботи
- проведення екзамену.

9. Схема нарахування балів

Поточний контроль, самостійна робота, індивідуальні завдання				Екзамен	Сума	
T1-T2	T3-T4	T5-T6	T7-T9	Контрольна робота, передбачена навчальним планом	Разом	
10	20	10	10	10	60	
					40	100

T1, T2 ... – теми.

Шкала оцінювання

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка	
	для чотирирівневої шкали оцінювання	для дворівневої шкали оцінювання
90 – 100	відмінно	зараховано
70-89	добре	
50-69	задовільно	
1-49	незадовільно	не зараховано

Критерії оцінювання навчальних досягнень

Оцінка в балах	Оцінка за національною шкалою	
Оцінка	Пояснення	
90 – 100	Відмінно	Теоретичний зміст курсу освоєний цілком, необхідні практичні навички роботи з освоєним матеріалом сформовані, всі навчальні завдання, які передбачені програмою навчання виконані в повному обсязі, відмінна робота без помилок або з однією незначною помилкою.
70 – 89	Добре	Теоретичний зміст курсу освоєний цілком, практичні навички роботи з освоєним матеріалом в основному сформовані, всі навчальні завдання, які передбачені програмою навчання виконані, якість виконання жодного з них не оцінено мінімальним числом балів, деякі види завдань виконані з помилками, робота з декількома незначними помилками, або з однією – двома значними помилками.
50 – 69	Задовільно	Теоретичний зміст курсу освоєний не повністю, але прогалини не носять істотного характеру, необхідні практичні навички роботи з освоєним матеріалом в основному сформовані, більшість передбачених програмою навчання навчальних завдань виконано, деякі з виконаних завдань, містять помилки, робота з трьома значними помилками.
1–49	Незадовільно	Теоретичний зміст курсу не освоєно, необхідні практичні навички роботи не сформовані, всі виконані навчальні завдання містять грубі помилки, додаткова самостійна робота над матеріалом курсу не приведе до значимого підвищення якості виконання навчальних завдань, робота, що потребує повної переробки

10. Рекомендована література

Основна література

1. Флах П. Машинное обучение. — М.: ДМК Пресс, 2015. — 400 с.
2. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. Учебник.—М.: Финансы и статистика, 2000—352 с.
3. Шитиков В.К., Мاستицкий С.Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R. — Тольятти, Лондон (2017). — 351с.
<https://github.com/ranalytics/data-mining>

Допоміжна література

1. Anil K. Jain, Richard C. Dubes Algorithms for Clustering Data, Prentice-Hall, Inc. Upper Saddle River, NJ, USA ,1988
2. S. Sra et al.. Optimization for Machine Hastie, T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. — 2nd ed. — Springer-Verlag, 2009. — 746 p.
3. Learning, MIT Press, 2011.

11. Посилання на інформаційні ресурси в Інтернеті, відео-лекції, інше методичне забезпечення

1. Вики-ресурс: <http://www.MachineLearning.ru>
2. Курс лекцій “Математические методы обучения по прецедентам:”
<http://www.ccas.ru/voron/teaching.html>
3. Курс «Машинное обучение»:
http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_%28%D0%BA%D1%83%D1%80%D1%81_%D0%BB%D0%B5%D0%BA%D1%86%D0%B8%D0%B9%2C_%D0%9A.%D0%92.%D0%92%D0%BE%D1%80%D0%BE%D0%BD%D1%86%D0%BE%D0%B2%29
4. Відео-лекції школи аналізу даних з машинного навчання
https://www.youtube.com/channel/UCKFojzto0n4Ab3CRQRZ2zYA/playlists?view=50&sort=dd&shelf_id=2