

Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Кафедра хімічного матеріалознавства

“ЗАТВЕРДЖУЮ”

Проректор з науково-педагогічної
роботи

“ _____ ” _____ 2019 р.

Робоча програма навчальної дисципліни

ХЕМОІНФОРМАТИКА І ХЕМОМЕТРІЯ

рівень вищої освіти магістр

галузь знань 10 Природничі науки

спеціальність 102 Хімія

освітня програма освітня-професійна програма “Хімія”

вид дисципліни за вибором

факультет хімічний

2019 / 2020 навчальний рік

Програму рекомендовано до затвердження вченою радою хімічного факультету

“ ____ ” _____ 2019 року, протокол № ____

РОЗРОБНИКИ ПРОГРАМИ:

Іванов В. В., д.х.н., професор кафедри хімічного матеріалознавства.

Пантелеймонов А. В., к.х.н., доцент кафедри хімічного матеріалознавства

Програму схвалено на засіданні кафедри хімічного матеріалознавства

Протокол від “ ____ ” _____ 2019 року № ____

Завідувач кафедри хімічного матеріалознавства

_____ Коробов О.І.
(підпис) (прізвище та ініціали)

Програму погоджено методичною комісією хімічного факультету

Протокол від “ ____ ” _____ 2019 року № ____

Голова методичної комісії хімічного факультету

_____ Єфімов П.В.
(підпис) (прізвище та ініціали)

ВСТУП

Програма навчальної дисципліни “Хемоінформатика і хеометрія” складена відповідно до освітньо-професійної (освітньо-наукової) програми підготовки магістра напрямку 102 – «Хімія»

1. Опис навчальної дисципліни

1.1. **Мета** викладання навчальної дисципліни – надати студентам, які навчаються за напрямом „Хімія” за ОКР магістр (спеціаліст) знань з використання хеометричних (статистичних та інших математичних) методів обробки масивів експериментальних даних, зокрема, результатів кількісного фізико-хімічного аналізу, квантово-хімічних методів розрахунку геометрії, електронної будови та властивостей складних багатоатомних систем.

1.2. Основні завдання вивчення дисципліни «Хемоінформатика і хеометрія» полягають у

- ознайомленні студентів із статистичними та математичними методами що направлені на розв’язання типових навчальних та наукових задач хімії;
- ознайомленні студентів із комп’ютерними технологіями, що направлені на обробку великих масивів структурно-хімічних даних;
- оволодінні програмними засобами для розв’язання прикладних хімічних задач;
- розвитку уміння роботи із доступними квантово-хімічними та статистичними пакетами програм.

1.3. Кількість кредитів 5

1.4. Загальна кількість годин 150

1.5. Характеристика навчальної дисципліни	
За вибором	
Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
1-й	1-й
Семестр	
1-й	19-й
Лекції	
16 год.	8 год.
Лабораторні заняття	
16 год.	8 год.
Самостійна робота	
118 год.	134 год.
Індивідуальні завдання	
Не передбачено	

1.6. Заплановані результати навчання. У результаті вивчення даного курсу студент повинен:

знати: засади хеометрії як міждисциплінарної науки для вилучення змістовної інформації з експериментальних даних та обробки даних хімічного експерименту.

вміти: використовувати комплекс хеометричних методів та розрахункових засобів для оцінювання параметрів хімічних систем.

2. Тематичний план навчальної дисципліни

Розділ 1. Лекційний матеріал

Тема 1. Хемометрія як міждисциплінарна наукова дисципліна. Хемометрія та хімія, хемометрія та прикладна статистика. Огляд завдань хемометрії та головних хемометричних методів.

Тема 2. Засади параметричної статистики. Розподіл похибок Гауса як граничний. Метод максимуму правдоподібності. Метод найменших квадратів (МНК). Розрахункова схема МНК. Некоректний характер задачі МНК та засоби регуляризації (Тихонівська регуляризація, сингулярний розклад як одна з реалізацій Тихонівської регуляризації). Статистичні властивості оцінок МНК. Аналіз адекватності моделей (глобальні та локальні критерії). Закон розподілу Лапласа як граничний. Метод найменших модулів (МНМ). Розрахункова схема МНМ.

Тема 3. Робастні оцінки. Різновиди робастних оцінок. М-оцінки Хьюбера. Розрахункова схема М-оцінювання. Робастні оцінки теорії нечітких множин (ТНМ). Розрахункова схема оцінювання за допомогою ТНМ.

Тема 4. Крос-валідація як засіб оцінки адекватності та верифікації моделей. Параметрична ідентифікація на основі крос-валідації.

Тема 5. Кластерний і дискримінантний аналіз. Приклади використання для вилучення змістовної хімічної інформації з великих масивів експериментальних даних.

Тема 6. Історичний огляд досліджень зв'язку біологічної активності та структури молекул. Методи розрахунків біологічної активності (БА) хімічних речовин як частина проблематики QSAR. Основні етапи оцінки БА. Загальний тезаурус типів БА.

Тема 7. Пакети прикладних програм орієнтовані на проблематику QSAR: Avogadro, HYPERCHEM, DRAGON, ACDLabs, PASS, тощо. Програми візуалізації. Візуалізація біомолекул.

Тема 8. Метод програмованих нейронних мереж. Типи нейронних мереж. Генетичні алгоритми. Застосування до проблем класифікації та кластеризації.

Тема 9. Аналіз багатовимірних даних. Метод головних компонент. Факторний аналіз багатоваріантних даних. Обертання факторів.

Тема 10. L_1 - та L_2 – регуляризація. Скорочення предикторного набору. Метод LASSO.

Розділ 2. Лабораторні заняття

Тема 11. Засади параметричної статистики. Розподіл похибок Гауса як граничний. Метод максимуму правдоподібності. Метод найменших квадратів (МНК). Розрахункова схема МНК. Некоректний характер задачі МНК та засоби регуляризації (Тихонівська регуляризація, сингулярний розклад як одна з реалізацій Тихонівської регуляризації). Статистичні властивості оцінок МНК. Аналіз адекватності моделей (глобальні та локальні критерії).

Тема 12. Закон розподілу Лапласа як граничний. Метод найменших модулів (МНМ). Розрахункова схема МНМ.

Тема 13. Робастні оцінки. Різновиди робастних оцінок. М-оцінки Хьюбера. Розрахункова схема М-оцінювання. Робастні оцінки теорії нечітких множин (ТНМ). Розрахункова схема оцінювання за допомогою ТНМ.

Тема 14. Крос-валідація як засіб оцінки адекватності та верифікації моделей. Параметрична ідентифікація на основі крос-валідації.

Тема 15. Розв'язання некоректно поставлених задач, що виникають при оцінці енергетичної неоднорідності поверхонь.

Тема 16. Кластерний і дискримінантний аналіз. Приклади використання для вилучення змістовної хімічної інформації з великих масивів експериментальних даних.

Тема 17. Знайомство з програмним пакетом MORAS. Оптимізація геометрії всевалентними методами. Порівняння розрахованих даних із експериментальними.

Тема 18. Використання нейронних мереж для прогнозу канцерогенної активності конденсованих вуглеводнів.

Тема 19. Факторний аналіз даних. Прогностичні моделі канцерогенної активності поліароматичних вуглеводнів.

Тема 20. Скорочення предикторного набору за допомогою методу LASSO. Регресійна модель канцерогенності.

3. Структура навчальної дисципліни

Назви розділів	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
Розділ 1. Лекційний матеріал												
Разом за розділом 1	76	16				60	76	8				68
Розділ 2. Лабораторні заняття												
Разом за розділом 2	74			16		58	74			8		66
Усього годин	150	16		16		118	150	8		8		134

4. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин	
		денне	заочне
11	Засади параметричної статистики. Розподіл похибок Гауса як граничний. Метод максимуму правдоподібності. Метод найменших квадратів (МНК). Розрахункова схема МНК. Некоректний характер задачі МНК та засоби регуляризації (Тихонівська регуляризація, сингулярний розклад як одна з реалізацій Тихонівської регуляризації). Статистичні властивості оцінок МНК. Аналіз адекватності моделей (глобальні та локальні критерії).	2	1
12	Закон розподілу Лапласа як граничний. Метод найменших модулів (МНМ). Розрахункова схема МНМ.	1	0.5
13	Робастні оцінки. Різновиди робастних оцінок. М-оцінки Хьюбера. Розрахункова схема М-оцінювання. Робастні оцінки теорії нечітких множин (ТНМ). Розрахункова схема оцінювання за допомогою ТНМ.	2	0.5
14	Крос-валідація як засіб оцінки адекватності та верифікації моделей. Параметрична ідентифікація на основі крос-валідації.	1	1
15	Розв'язання некоректно поставлених задач, що виникають при оцінці енергетичної неоднорідності поверхонь.	2	1
16	Кластерний і дискримінантний аналіз. Приклади використання для вилучення змістовної хімічної інформації з великих масивів експериментальних даних.	1	1
17	Знайомство з програмним пакетом MORAS. Оптимізація геометрії всевалентними методами. Порівняння розрахованих даних із експериментальними.	2	1
18	Використання нейронних мереж для прогнозу канцерогенної активності конденсованих вуглеводнів.	2	1

19	Факторний аналіз даних. Прогностичні моделі канцерогенної активності поліароматичних вуглеводнів.	2	0.5
20	Скорочення предикторного набору за допомогою методу LASSO. Регресійна модель канцерогенності.	1	0.5
	Разом	16	8

5. Самостійна робота

Назва теми	Кількість годин	
	денне	заочне
Тема 1. Хемометрія як міждисциплінарна наукова дисципліна. Хемометрія та хімія, хемометрія та прикладна статистика. Огляд завдань хемометрії та головних хемометричних методів.	5	3
Тема 2. Засади параметричної статистики. Розподіл похибок Гауса як граничний. Метод максимуму правдоподібності. Метод найменших квадратів (МНК). Розрахункова схема МНК. Некоректний характер задачі МНК та засоби регуляризації (Тихонівська регуляризація, сингулярний розклад як одна з реалізацій Тихонівської регуляризації). Статистичні властивості оцінок МНК. Аналіз адекватності моделей (глобальні та локальні критерії). Закон розподілу Лапласа як граничний. Метод найменших модулів (МНМ). Розрахункова схема МНМ.	5	7
Тема 3. Робастні оцінки. Різновиди робастних оцінок. М-оцінки Хьюбера. Розрахункова схема М-оцінювання. Робастні оцінки теорії нечітких множин (ТНМ). Розрахункова схема оцінювання за допомогою ТНМ.	7	5
Тема 4. Крос-валідація як засіб оцінки адекватності та верифікації моделей. Параметрична ідентифікація на основі крос-валідації.	5	7
Тема 5. Кластерний і дискримінантний аналіз. Приклади використання для вилучення змістовної хімічної інформації з великих масивів експериментальних даних.	7	5
Тема 6. Історичний огляд досліджень зв'язку біологічної активності та структури молекул. Методи розрахунків біологічної активності (БА) хімічних речовин як частина проблематики QSAR. Основні етапи оцінки БА. Загальний тезаурус типів БА.	6	8
Тема 7. Пакети прикладних програм орієнтовані на проблематику QSAR: Avogadro, HYPERCHEM, DRAGON, ACDLabs, PASS, тощо. Програми візуалізації. Візуалізація біомолекул.	6	8
Тема 8. Метод програмованих нейронних мереж. Типи нейронних мереж. Генетичні алгоритми. Застосування до проблем класифікації та кластеризації.	6	8
Тема 9. Аналіз багатовимірних даних. Метод головних компонент. Факторний аналіз багатоваріантних даних. Обертання факторів.	6	8
Тема 10. L_1 - та L_2 – регуляризація. Скорочення предикторного набору. Метод LASSO.	7	9
Разом	60	68

6. Індивідуальні завдання

Не передбачено навчальним планом.

7. Методи контролю

Перевірка результатів лабораторних робіт (звітів). Виконання контрольних робіт. Семестровий екзамен (письмова робота).

8. Схема нарахування балів

Для допуску до семестрового екзамену студенти мають виконати усі лабораторні роботи.

Поточний контроль, самостійна робота, індивідуальні завдання					Екзамен	Сума
Розділ 1	Розділ 2	Контрольна робота, передбачена навчальним планом	Індивідуальне завдання	Разом		
T1–T10	T11 – T20					
–	60	–	–	60	40	100

Шкала оцінювання

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

Критерії оцінювання навчальних досягнень студентів з дисципліни

– «відмінно» – студент міцно засвоїв теоретичний матеріал, всебічно орієнтується в усіх основних темах курсу; здатен ефективно розв'язати поставлену проблему хімії та аналізувати одержані результати; відповідь надає обґрунтовано та логічно; вільно володіє основними програмними пакетами для вирішення комплексних завдань.

– «добре» – студент добре засвоїв теоретичний матеріал та орієнтується в усіх основних темах курсу; здатен розв'язати поставлену проблему, але не впевнено та обґрунтовано аналізує одержані результати; володіє основними програмними пакетами для вирішення навчальних завдань.

– «задовільно» – студент в основному засвоїв теоретичний матеріал, але орієнтується не в усіх основних темах курсу; здатен вирішувати типові навчальні завдання, але відповідь надає не повну; невпевнено володіє програмними пакетами для вирішення навчальних завдань.

– «незадовільно» – студент не засвоїв теоретичний матеріал, не орієнтується у більшості основних тем курсу; не здатен вирішувати типові навчальні завдання; не володіє основними програмними пакетами для вирішення навчальних завдань.

9. Рекомендована література

Основна література

1. Холин Ю.В. Количественный физико-химический анализ комплексообразования в растворах и на поверхности химически модифицированных кремнеземов: содержательные модели, математические методы и их приложения. – Харьков: *Фолио*, 2000. – 288 с.
2. Холин Ю.В., Никитина Н.А., Пантелеймонов А.В., Решетняк Е.А., Бугаевский А.А., Логинова Л.П. Метрологические характеристики методик обнаружения с бинарным откликом. – Харьков: *Тимченко*, 2008. – 128 с.

3. Иванов В. В., Слета Л. А.. Квантовая химия.– Харьков: “Фоліо”, 2007. – 443 с.
4. Иванов В.В., Слета Л.А., Расчетные методы прогноза биологической активности органических соединений. – Харьков: *ЧП Азамаев В.Р.*, 2003.– 76 с.
5. Орлов В.Д., Липсон В.В., Иванов В.В. Медицинская химия. – Харьков: “Фоліо”, 2005. – 461 с.

Допоміжна література

1. Huber P. Robust statistical procedures, in: CBMS-NSF Regional conf. series in appl. mathematics, SIAM, Philadelphia, 1996.
2. Родионова О.Е. Хемометрика: достижения и перспективы / О.Е. Родионова, А.Л. Померанцев // Успехи химии. – 2006. – Т. 75, № 4. – С. 302-321.
3. Wold S. Chemometrics: what do we mean with it, and what do we want from it? // Chemom. Intell. Lab. Syst. – 1995. – V. 30, No 1. – P. 109-115.
4. Massart D.L., Vandeginste B.G.M., Buydens L.M.C., de Jong S., Lewi P.J., Smeyers-Verbeke J. Handbook of Chemometrics and Qualimetrics: Part A. Amsterdam: Elsevier, 1997. – 886 p.
5. R.G. Brereton. Chemometrics: Data Analysis for the Laboratory and Chemical Plant , Wiley, Chichester, January 2003, 489 p.
6. Нейман Ю. Вводный курс теории вероятностей и математической статистики. – М.: Наука, 1968. – 448 с.
7. Вершинин В.И., Дерендяев Б.Г., Лебедев К.С. Компьютерная идентификация органических соединений. – М.: Академкнига, 2002. – 197 с.
8. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: В 2-х кн.
9. Young D.C. Computational Chemistry, Wiley Interscience, New York, 2001. 370 p.
10. Станкевич М.И., Станкевич И.В., Зефирова Н.С., Топологические индексы в органической химии // Успехи химии.–1988.–Т. 57, № 3.– С.337-365.
11. Раевский О.А., Дескрипторы молекулярной структуры в компьютерном дизайне биологически активных веществ // Успехи химии.–1999.–Т. 68, №6. – С. 555-575.
12. Факторный, дискриминантный и кластерный анализ, М.”Финансы и статистика”, 1989, 213 с.

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

Інформаційні ресурси

1. Файл-сервер хімічного факультету ХНУ імені В.Н. Каразіна:
<http://www-chemistry.univer.kharkov.ua/uk>
2. Сайт кафедри хімічного матеріалознавства
<http://www-chemo.univer.kharkov.ua/>