

Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Кафедра хімічного матеріалознавства

“ЗАТВЕРДЖУЮ”

Перший проректор

“ _____ ” _____ 20__ р.

ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Хемоінформатика і хеометрія

напря́м 040101 – «Хімія»

спеціальність 7.04010101– «Хімія», 8.04010101 – «Хімія»

хімічний факультет

2015 / 2016 навчальний рік

Програму обговорено та рекомендовано до затвердження Вченою радою хімічного факультету
27 серпня 2015 року, протокол № 7

Зі змінами та доповненнями затверджено Вченою радою хімічного факультету
25 вересня 2015 року, протокол № 8

РОЗРОБНИКИ ПРОГРАМИ:

Холін Юрій Валентинович, д.х.н., професор, в.о. завідувача кафедри хімічного матеріалознавства,

Іванов Володимир Венедиктович, д.х.н., професор, професор кафедри хімічного матеріалознавства,

Пантелеймонов Антон Віталійович, к.х.н., доцент, доцент кафедри хімічного матеріалознавства

Програму схвалено на засіданні кафедри хімічного матеріалознавства;
26 серпня 2015 року, протокол № 1

Зі змінами та доповненнями схвалено на засіданні кафедри хімічного матеріалознавства;
21 вересня 2015 року, протокол № 2

В.о. завідувача кафедри
хімічного матеріалознавства

Холін Ю.В.

Програму погоджено методичною комісією хімічного факультету
24 вересня 2015 року, протокол № 2

Голова методичної комісії хімічного факультету

Юрченко О.І.

ВСТУП

Програма навчальної дисципліни “Хемоінформатика і хеометрія” складена відповідно до освітньо-професійної (освітньо-наукової) програми підготовки магістра (спеціаліста) напряму 6.040101 – «Хімія»

Предметом вивчення навчальної дисципліни є Хемоінформатика як міждисциплінарна наукова концепція на границі хімії та інформатики та основи застосування хеометричних методів у хімії

Програма навчальної дисципліни складається з таких розділів:

1. Мета та завдання навчальної дисципліни.
2. Опис навчальної дисципліни.
3. Виклад змісту навчальної дисципліни.
4. Структура навчальної дисципліни.
5. Теми лабораторних занять.
6. Самостійна робота.
7. Індивідуальні завдання.
8. Методи навчання.
9. Методи контролю.
10. Розподіл балів, які отримують студенти.
11. Методичне забезпечення.

1. Мета та завдання навчальної дисципліни

1.1. Метою викладання навчальної дисципліни є надання 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. Основними завданнями вивчення дисципліни є

- знайомство студентів із комп’ютерними технологіями для розв’язання типових навчальних та наукових задач;
- оволодіння програмними засобами для розв’язання прикладних хімічних задач.

1.3. Згідно з вимогами освітньо-професійної (освітньо-наукової) програми студенти повинні досягти таких результатів навчання:

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. Опис навчальної дисципліни

Найменування показника	Галузь знань (предметна область), напрям, спеціальність, рівень вищої освіти / освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни	
		денна форма навчання	заочна форма навчання
Кількість кредитів 7	Галузь знань (предметна область): <u>0401 «Природничі науки»</u> Напрямок: <u>040101 «Хімія»</u>	за вибором	
Індивідуальне науково-дослідне завдання	Спеціальність: <u>7.04010101 – «Хімія»</u> <u>8.04010101 – «Хімія»</u>	Рік підготовки	
Загальна кількість годин 210		5-й	5-й
Тижневих годин для денної форми навчання: аудиторних – 72 самостійної роботи студента – 138	Рівень вищої освіти (освітньо-кваліфікаційний рівень): <u>спеціаліст,</u> <u>магістр</u>	Семестр	
		9-й	9-й
		Лекції	
		36 год.	4 год.
		Практичні, семінарські	
		год.	8 год.
		Лабораторні	
		36 год.	год.
		Самостійна робота	
		138 год.	108 год.
Індивідуальні завдання:			
год.			
Вид контролю:			
екзамен	екзамен		

Співвідношення кількості годин аудиторних занять до самостійної і індивідуальної роботи становить (%):

для денної форми навчання – 150%

для заочної форми навчання – 11%

3. Виклад змісту навчальної дисципліни

Модуль 1. Лекції (частина 1)

Тема 1. Хемометрія як міждисциплінарна наукова дисципліна. Хемометрія та хімія, хемометрія та прикладна статистика.

Тема 2. Огляд завдань хемометрії та головних хемометричних методів.

Тема 3. 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 4. Закон розподілу Лапласа як граничний. Метод найменших модулів (МНМ). Розрахункова схема МНМ.

Тема 5. 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 6. Крос-валідація як засіб оцінки адекватності та верифікації моделей. Параметрична ідентифікація на основі крос-валідації.

Тема 7. Розв'язання некоректно поставлених задач, що виникають при оцінці енергетичної неоднорідності поверхонь.

Тема 8. Кластерний і дискримінантний аналіз. Приклади використання для вилучення змістовної хімічної інформації з великих масивів експериментальних даних.

Тема 9. Кореляційний аналіз. Кореляційні рівняння в хімії.

Тема 10. Проблема розробки математичних моделей, що пов'язують структуру речовини із її властивостями. Основні підходи. Квантова хімія як метод прогнозу та цілеспрямованого створення систем із заданими властивостями.

Тема 11. Систематика квантовохімічних методів. Напівемпіричні та неемпіричні (*ab initio*) методи. Метод Гартрі-Фока. Всевалентні наближення – CNDO, MNDO, MINDO, AM1, PM3, RM1, PM6. Їх параметризація. Знайомство з програмними пакетами HyperChem. MOPAC.

Тема 12. Розрахунки електронної будови та геометрії біологічно активних сполук. Розрахунки фізико-хімічних властивостей молекул. Фуллерен, наносистеми, вуглецеві нанотрубки, графен.

Тема 13. 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 14. Пакети прикладних програм орієнтовані на проблематику QSAR: Avogadro, HYPERCHEM, DRAGON, ACDLabs, PASS, тощо. Програми візуалізації. Візуалізація біомолекул.

Тема 15. Логіко-комбінаторний метод у конструюванні лікарських препаратів. Байєсові оцінки ймовірності.

Тема 16. Метод програмованих нейронних мереж. Типи нейронних мереж. Генетичні алгоритми. Застосування до проблем класифікації та кластеризації.

Тема 17. Аналіз багатовимірних даних. Метод головних компонент. Факторний аналіз багатоваріантних даних. Обертання факторів.

Тема 18. L_1 - та L_2 – регуляризація. Скорочення предикторного набору. Метод LASSO.

Тема 19. Дискримінаційний аналіз та логістична регресія в проблемі побудови прогностичних моделей біологічної активності. Канцерогенна активність полі циклічних ароматичних сполук.

Модуль 2. Лабораторні заняття

Тема 20. 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 21. Закон розподілу Лапласа як граничний. Метод найменших модулів (МНМ). Розрахункова схема МНМ.

Тема 22. 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 23. Крос-валідація як засіб оцінки адекватності та верифікації моделей. Параметрична ідентифікація на основі крос-валідації.

Тема 24. Розв'язання некоректно поставлених задач, що виникають при оцінці енергетичної неоднорідності поверхонь.

Тема 25. Кластерний і дискримінантний аналіз. Приклади використання для вилучення змістовної хімічної інформації з великих масивів експериментальних даних.

Тема 26. Кореляційний аналіз. Кореляційні рівняння в хімії.

Тема 27. Знайомство з програмним пакетом MORAC. Оптимізація геометрії всевалентними методами. Порівняння розрахованих даних із експериментальними.

Тема 28. «Сборка» наносистем за допомогою візуалізуючих програм. Розрахунки електронної будови вуглецевих нанотрубок.

Тема 29. Розрахунки біологічної активності за допомогою *on-line* ресурсу PASS.

Тема 30. Побудова прогностичних моделей за логіко-комбінаторним методом.

Тема 31. Використання нейронних мереж для прогнозу канцерогенної активності конденсованих вуглеводнів.

Тема 32. Факторний аналіз даних. Прогностичні моделі канцерогенної активності поліароматичних вуглеводнів.

Тема 33. Скорочення предикторного набору за допомогою методу LASSO. Регресійна модель канцерогенності.

4. 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
Розділ 1. Виклад теоретичного матеріалу												
Тема 1.	5	2				3						
Тема 2.	5	2				3						
Тема 3.	5	2				3						
Тема 4.	6	2				4						
Тема 5.	6	2				4						
Тема 6.	6	2				4						
Тема 7.	6	2				4						
Тема 8.	6	2				4						
Тема 9.	6	2				4						
Тема 10.	5	1				4						
Тема 11.	6	2				4						
Тема 12.	6	2				4						
Тема 13.	6	2				4						
Тема 14.	6	2				4						
Тема 15.	6	2				4						
Тема 16.	5	2				3						
Тема 17.	5	2				3						
Тема 18.	5	2				3						
Тема 19.	4	1				3						
Разом за розділом 1	105	36				69						
Розділ 2. Лабораторні заняття												
Тема 20.	7			2		5						
Тема 21.	7			2		5						
Тема 22.	7			2		5						

Тема 23.	8		3	5								
Тема 24.	8		3	5								
Тема 25.	8		3	5								
Тема 26.	8		3	5								
Тема 27.	8		3	5								
Тема 28.	8		3	5								
Тема 29.	8		2	5								
Тема 30.	7		3	5								
Тема 31.	7		2	5								
Тема 32.	7		2	5								
Тема 33.	7		3	4								
Разом за розділом 2	105		36	69								
Усього годин	210	36	36	138								

5. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
Модуль 2		
20	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
21	Закон розподілу Лапласа як граничний. Метод найменших модулів (МНМ). Розрахункова схема МНМ.	2
22	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
23	Крос-валідація як засіб оцінки адекватності та верифікації моделей. Параметрична ідентифікація на основі крос-валідації.	3
24	Розв'язання некоректно поставлених задач, що виникають при оцінці енергетичної неоднорідності поверхонь.	3
25	Кластерний і дискримінантний аналіз. Приклади використання для вилучення змістовної хімічної інформації з великих масивів експериментальних даних.	3
26	Кореляційний аналіз. Кореляційні рівняння в хімії.	3
27	Знайомство з програмним пакетом MORAS. Оптимізація геометрії всевалентними методами. Порівняння розрахованих даних із експериментальними.	3
28	«Сборка» наносистем за допомогою візуалізуючих програм. Розрахунки електронної будови вуглецевих нанотрубок.	3
29	Розрахунки біологічної активності за допомогою <i>on-line</i> ресурсу PASS .	2
30	Побудова прогностичних моделей за логіко-комбінаторним методом.	3
31	Використання нейронних мереж для прогнозу канцерогенної	2

	активності конденсованих вуглеводнів.	
32	Факторний аналіз даних. Прогностичні моделі канцерогенної активності полі циклічних ароматичних вуглеводнів.	2
33	Скорочення предикторного набору за допомогою методу LASSO. Регресійна модель канцерогенності.	3
	Усього годин	36

6. Самостійна робота

Назва теми	Кількість годин	Форма контролю
Тема 1. Хемометрія як міждисциплінарна наукова дисципліна. Хемометрія та хімія, хемометрія та прикладна статистика.	3	Поточний контроль на лекціях.
Тема 2. Огляд завдань хемометрії та головних хемометричних методів.	3	
Тема 3. 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).	3	Виконання контрольних робіт. Семестровий екзамен (письмова робота).
Тема 4. Закон розподілу Лапласа як граничний. Метод найменших модулів (МНМ). Розрахункова схема МНМ.	4	
Тема 5. Робастні оцінки. Різновиди робастних оцінок. М-оцінки Хьюбера. Розрахункова схема М-оцінювання. Робастні оцінки теорії нечітких множин (ТНМ). Розрахункова схема оцінювання за допомогою ТНМ.	4	
Тема 6. Крос-валідація як засіб оцінки адекватності та верифікації моделей. Параметрична ідентифікація на основі крос-валідації.	4	
Тема 7. Розв'язання некоректно поставлених задач, що виникають при оцінці енергетичної неоднорідності поверхонь.	4	
Тема 8. Кластерний і дискримінантний аналіз. Приклади використання для вилучення змістовної хімічної інформації з великих масивів експериментальних даних.	4	
Тема 9. Кореляційний аналіз. Кореляційні рівняння в хімії.	4	
Тема 10. Проблема розробки математичних моделей, що пов'язують структуру речовини із її властивостями. Основні підходи. Квантова хімія як метод прогнозу та цілеспрямованого створення систем із заданими властивостями.	4	
Тема 11. Систематика квантовохімічних методів. Напівемпіричні та неемпіричні (<i>ab initio</i>) методи. Метод Гюккеля та ППП. Параметризація ППП. Метод Гартрі-Фока. Всевалентні наближення – CNDO, MNDO, MINDO, AM1, PM3, RM1, PM6. Молекулярна механіка. Знайомство з програмними пакетами HyperChem. MORAC.	4	
Тема 12. π -електронна будова та властивості полімерних π -спряжених систем поліацетілену, кумулени, конденсовані вуглеводні. інтерпретації фізико-хімічних властивостей	4	

молекул. Наносистеми. Фуллерен, Вуглецеві нанотрубки, графен.		
Тема 13. 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.	4	
Тема 14. Пакети прикладних програм орієнтовані на проблематику QSAR: Avogadro, HYPERCHEM, DRAGON, ACDLabs, PASS, тощо. Програми візуалізації. Візуалізація біомолекул.	4	
Тема 15. Логіко-комбінаторний метод у конструюванні лікарських препаратів. Байесові оцінки ймовірності.	4	
Тема 16. Метод програмованих нейронних мереж. Генетичні алгоритми. Застосування до проблем класифікації та кластеризації.	3	
Тема 17. Аналіз багатовимірних даних. Метод головних компонент. Факторний аналіз багатоваріантних даних. Обертання факторів.	3	
Тема 18. L_1 - та L_2 – регуляризація. Скорочення предикторного набору. Метод LASSO.	3	
Тема 19. Дискримінаційний аналіз та логістична регресія в проблемі побудови прогностичних моделей біологічної активності. Канцерогенна активність полі циклічних ароматичних сполук.	3	
Тема 20. 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).	5	
Тема 21. Закон розподілу Лапласа як граничний. Метод найменших модулів (МНМ). Розрахункова схема МНМ.	5	
Тема 22. Робастні оцінки. Різновиди робастних оцінок. М-оцінки Хьюбера. Розрахункова схема М-оцінювання. Робастні оцінки теорії нечітких множин (ТНМ). Розрахункова схема оцінювання за допомогою ТНМ.	5	
Тема 23. Крос-валідація як засіб оцінки адекватності та верифікації моделей. Параметрична ідентифікація на основі крос-валідації.	5	
Тема 24. Розв'язання некоректно поставлених задач, що виникають при оцінці енергетичної неоднорідності поверхонь.	5	
Тема 25. Кластерний і дискримінантний аналіз. Приклади використання для вилучення змістовної хімічної інформації з великих масивів експериментальних даних.	5	
Тема 26. Кореляційний аналіз. Кореляційні рівняння в хімії.	5	
Тема 27. Знайомство з програмним пакетом MORAC. Оптимізація геометрії всевалентними методами. Порівняння розрахованих даних із експериментальними.	5	

Тема 28. «Сборка» наносистем за допомогою візуалізуючих програм. Розрахунки електронної будови вуглецевих нанотрубок.	5	
Тема 29. Розрахунки біологічної активності за допомогою <i>on-line</i> ресурсу PASS .	5	
Тема 30. Побудова прогностичних моделей за логіко-комбінаторним методом.	5	
Тема 31. Використання нейронних мереж для прогнозу канцерогенної активності конденсованих вуглеводнів.	5	
Тема 32. Факторний аналіз даних.	5	
Тема 33. Скорочення предикторного набору за допомогою методу LASSO. Регресійна модель канцерогенності.	4	

7. Індивідуальні завдання

Не передбачено навчальним планом.

8. Методи навчання

Лекції, лабораторні заняття, самостійна робота.

9. Методи контролю

Поточний контроль на лекціях. Виконання контрольних робіт. Семестровий екзамен (письмова робота).

10. Розподіл балів, які отримують студенти

Поточне тестування та самостійна робота	Підсумковий семестровий контроль (екзамен)	Сума
Модуль 2		
60		
	40	100

Шкала оцінювання

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка за національною шкалою
90-100	відмінно
70-89	добре
50-69	задовільно
1-49	незадовільно

11. Рекомендоване методичне забезпечення

- Програма навчальної дисципліни.
- Підручники.
- Навчальні посібники.
- Описи лабораторних робіт.
- Набори контрольних питань для поточного контролю знань.
- Завдання для контрольних робіт.
- Екзаменаційні білети.

Базова 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: “*Фолио*”, 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: “*Фолио*”, 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.

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/>