

Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Кафедра хімічного матеріалознавства

“ЗАТВЕРДЖУЮ”

Проректор з науково-педагогічної
роботи

“ _____ ” _____ 2018 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

2018 / 2019 навчальний рік

Програму рекомендовано до затвердження вченою радою хімічного факультету

“ 31 ” серпня 2018 року, протокол № 7

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 “31” серпня 2018 року № 1

Завідувач кафедри хімічного матеріалознавства

(підпис)

Коробов О.І.

(прізвище та ініціали)

Програму погоджено методичною комісією хімічного факультету

Протокол від “31” серпня 2018 року № 1

Голова методичної комісії хімічного факультету

(підпис)

Єфімов П.В.

(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 7

1.4. Загальна кількість годин 210

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	
32 год.	10 год.
Лабораторні заняття	
32 год.	10 год.
Самостійна робота	
146 год.	190 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.

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.

Розділ 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 МОРАС. 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.

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	105	32				73	105	10				95
Розділ 2. Лабораторні заняття												
Разом за розділом 2	105			32		73	105			10		95
Усього годин	210	32		32		146	210	10		10		190

4. Теми лабораторних занять

№ з/п	Назва теми	Кількість Годин	
		денне	з. в.
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).	3	1
21	Закон розподілу Лапласа як граничний. Метод найменших модулів (МНМ). Розрахункова схема МНМ.	3	0.5
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	0.5
23	Крос-валідація як засіб оцінки адекватності та верифікації моделей. Параметрична ідентифікація на основі крос-валідації.	2	1

24	Розв'язання некоректно поставлених задач, що виникають при оцінці енергетичної неоднорідності поверхонь.	2	1
25	Кластерний і дискримінантний аналіз. Приклади використання для вилучення змістовної хімічної інформації з великих масивів експериментальних даних.	2	1
26	Кореляційний аналіз. Кореляційні рівняння в хімії.	2	1
27	Знайомство з програмним пакетом MORAC. Оптимізація геометрії всевалентними методами. Порівняння розрахованих даних із експериментальними.	3	1
28	«Сборка» наносистем за допомогою візуалізуючих програм. Розрахунки електронної будови вуглецевих нанотрубок.	2	0.5
29	Розрахунки біологічної активності за допомогою <i>on-line</i> ресурсу PASS.	2	0.5
30	Побудова прогностичних моделей за логіко-комбінаторним методом.	2	0.5
31	Використання нейронних мереж для прогнозу канцерогенної активності конденсованих вуглеводнів.	3	0.5
32	Факторний аналіз даних. Прогностичні моделі канцерогенної активності полі циклічних ароматичних вуглеводнів.	2	0.5
33	Скорочення предикторного набору за допомогою методу LASSO. Регресійна модель канцерогенності.	2	0.5
Усього годин		32	10

5. Самостійна робота

Назва теми	Кількість годин	
	денне	заочне
Тема 1. Хемометрія як міждисциплінарна наукова дисципліна. Хемометрія та хімія, хемометрія та прикладна статистика.	4	4
Тема 2. Огляд завдань хемометрії та головних хемометричних методів.	4	8
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).	6	6
Тема 4. Закон розподілу Лапласа як граничний. Метод найменших модулів (МНМ). Розрахункова схема МНМ.	4	8
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.	6	6
Тема 6. Крос-валідація як засіб оцінки адекватності та верифікації моделей. Параметрична ідентифікація на основі крос-валідації.	4	8
Тема 7. Розв'язання некоректно поставлених задач, що виникають при оцінці енергетичної неоднорідності поверхонь.	4	8
Тема 8. Кластерний і дискримінантний аналіз. Приклади використання для вилучення змістовної хімічної інформації з великих масивів експериментальних даних.	4	8
Тема 9. Кореляційний аналіз. Кореляційні рівняння в хімії.	4	8
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.	6	10
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,	6	10

MINDO, AM1, PM3, RM1, PM6. Молекулярна механіка. Знайомство з програмними пакетами HyperChem. MOPAC.		
Тема 12. π -електронна будова та властивості полімерних π -спряжених систем поліацетілену, кумулени, конденсовані вуглеводні. інтерпретації фізико-хімічних властивостей молекул. Наносистеми. Фуллерен, Вуглецеві нанотрубки, графен.	6	10
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.	6	10
Тема 14. Пакети прикладних програм орієнтовані на проблематику QSAR: Avogadro, HYPERCHEM, DRAGON, ACDLabs, PASS, тощо. Програми візуалізації. Візуалізація біомолекул.	6	10
Тема 15. Логіко-комбінаторний метод у конструюванні лікарських препаратів. Байєсові оцінки ймовірності.	4	4
Тема 16. Метод програмованих нейронних мереж. Генетичні алгоритми. Застосування до проблем класифікації та кластеризації.	4	4
Тема 17. Аналіз багатовимірних даних. Метод головних компонент. Факторний аналіз багатоваріантних даних. Обертання факторів.	4	4
Тема 18. L_1 - та L_2 – регуляризація. Скорочення предикторного набору. Метод LASSO.	4	4
Тема 19. Дискримінаційний аналіз та логістична регресія в проблемі побудови прогностичних моделей біологічної активності. Канцерогенна активність поліциклічних ароматичних сполук.	4	4
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).	4	4
Тема 21. Закон розподілу Лапласа як граничний. Метод найменших модулів (МНМ). Розрахункова схема МНМ.	4	4
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.	4	4
Тема 23. Крос-валідація як засіб оцінки адекватності та верифікації моделей. Параметрична ідентифікація на основі крос-валідації.	4	4
Тема 24. Розв'язання некоректно поставлених задач, що виникають при оцінці енергетичної неоднорідності поверхонь.	4	4
Тема 25. Кластерний і дискримінантний аналіз. Приклади використання для вилучення змістовної хімічної інформації з великих масивів експериментальних даних.	4	4
Тема 26. Кореляційний аналіз. Кореляційні рівняння в хімії.	4	4
Тема 27. Знайомство з програмним пакетом MOPAC. Оптимізація геометрії всевалентними методами. Порівняння розрахованих даних із експериментальними.	4	4
Тема 28. «Сборка» наносистем за допомогою візуалізуючих програм. Розрахунки електронної будови вуглецевих нанотрубок.	4	4
Тема 29. Розрахунки біологічної активності за допомогою <i>on-line</i> ресурсу PASS .	4	4
Тема 30. Побудова прогностичних моделей за логіко-комбінаторним методом.	4	4
Тема 31. Використання нейронних мереж для прогнозу канцерогенної активності конденсованих вуглеводнів.	4	4

Тема 32. Факторний аналіз даних.	4	4
Тема 33. Скорочення предикторного набору за допомогою методу LASSO. Регресійна модель канцерогенності.	4	4
Разом	146	190

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			Разом	Екзамен	Сума
Розділ 1	Розділ 2				
T1–T20	T20 – T26	T27-T33			
–	30	30	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

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: “*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., Зефиоров 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/>