

# Облікова картка дисертації

## I. Загальні відомості

Державний обліковий номер: 0822U100191

Особливі позначки: відкрита

Дата реєстрації: 14-01-2022

Статус: Захищена

Реквізити наказу МОН / наказу закладу:



## II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Бердник Михайло Ігорович

2. Berdnyk Mykhailo I.

Кваліфікація:

Ідентифікатор ORCID ID: Не застосовується

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 102

Назва наукової спеціальності: Хімія

Галузь / галузі знань:

Освітньо-наукова програма зі спеціальності: Не застосовується

Дата захисту: 28-12-2021

Спеціальність за освітою: хімія

Місце роботи здобувача:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

### **III. Відомості про організацію, де відбувся захист**

**Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради):** ДФ 64.051.041

**Повне найменування юридичної особи:** Харківський національний університет імені В. Н. Каразіна

**Код за ЄДРПОУ:** 02071205

**Місцезнаходження:** майдан Свободи, буд. 4, м. Харків, Харківський р-н., Харківська обл., 61022, Україна

**Форма власності:**

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:** Не застосовується

### **IV. Відомості про підприємство, установу, організацію, в якій було виконано дисертацію**

**Повне найменування юридичної особи:** Харківський національний університет імені В. Н. Каразіна

**Код за ЄДРПОУ:** 02071205

**Місцезнаходження:** майдан Свободи, буд. 4, м. Харків, Харківський р-н., Харківська обл., 61022, Україна

**Форма власності:**

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:** Не застосовується

### **V. Відомості про дисертацію**

**Мова дисертації:**

**Коди тематичних рубрик:** 13.31.21.15, 31.21

**Тема дисертації:**

1. Метод L1 регуляризації для опису фізико-хімічних властивостей молекул
2. L1-regularization method for the description of the physical and chemical properties of molecules

**Реферат:**

1. Роботу присвячено дослідженню можливостей використання L1-регуляризації в побудові хеометричних моделей «структура-активність» і квантовохімічних розрахунках. Для виконання завдань дисертації розроблено оригінальний комплекс програм, що реалізують різні статистичні (хеометричні) підходи до побудови регресійних моделей й аналізу їх прогностичної здатності. Також створено комплекс квантовохімічних програм, у яких L1-регуляризація використовується для побудови хвильових функцій методів, що ураховують електронну кореляцію. Зокрема, у дисертаційній роботі розглядалося використання L1-регуляризації для побудови лінійних емпіричних моделей опису різних фізико-хімічних параметрів молекул. Спираючись на досліджені вибірки молекул, було показано, що з використанням L1-регуляризації завжди можна сформулювати послідовний (упорядкований) набір дескрипторів. Систематично додаючи дескриптори з цього набору до моделей лінійної регресії або штучних нейронних мереж, можна отримати регресійні моделі з послідовно зростаючими величинами критеріїв валідації. Оскільки після ранжування дескрипторного набору обрані предиктори можуть використовуватися в різних підходах до побудови

лінійної регресії, нами було проведено відповідне дослідження якості цих альтернативних моделей. Було показано, що той чи інший метод може мати кращі прогностичні властивості відповідно до критеріїв зовнішньої або внутрішньої валідації. Показано, що методом штучних нейронних мереж з використанням впорядкованого дескрипторного набору, який був отриманий методом L1-регуляризації, також може бути зроблено якісні прогнози властивостей речовини. Також було проведено співставлення отриманих рівнянь лінійної регресії з альтернативними підходами, що працюють із нескороченими (неоптимізованими) дескрипторними наборами. У вивчених прикладах L1-регуляризація дозволила сформулювати компактні одно-, двух- або трьох- параметричні моделі, які здатні задовільно описати набір даних. Також, моделі отримані з попереднім відбором із використанням LARS-LASSO виявились кращими, ніж результати розрахунків PLS та PCR. Певну увагу в дисертації приділено методам валідації й оцінкам якості регресійних рівнянь. З цією метою було використано модельну задачу, у яку вносилися похибки як в залежну, так і в незалежну змінні, при цьому розглядався найпростіший випадок – регресія з однією незалежною змінною. З використанням модельної задачі було продемонстровано, що випадкове одиничне розбиття вибірки на тренувальну та тестову не є інформативним. Отже, для адекватної оцінки регресійного рівняння, а також дослідження якості вхідних даних у цілому, необхідно створювати та вивчати якомога більше розбивань на тренувальну й тестову вибірку. Також було досліджено відомі, запропоновані на сьогодні, критерії валідації. Встановлено, що для даних із вираженим розкидом типовою картиною є зворотна (суттєво нелінійна) залежність критеріїв зовнішньої валідації від внутрішньої. Інша тісно пов'язана із побудовою статистичних моделей проблема це побудова класифікаційної функції. З цією метою в роботі використано L1-регуляризований розрахунок логістичної регресії. На досліджених задачах побудови класифікаційних функцій показано, що з використанням L1-регуляризованої логістичної регресії можна досягнути конкурентно-спроможних результатів класифікації до результатів, отриманих з використанням інших, більш складних у розрахунковому сенсі, методів. Використання спеціального L1-регуляризованого алгоритму дало можливість отримати досить прості класифікаційні рівняння, які є інтерпретувемими. Також отримані рівняння логістичної регресії є однозначними й відтворюваними. Показано, що метод L1-регуляризації може бути використаний і в квантовій хімії. За допомогою процедури L1-регуляризації можливо створення впорядкованого (ранжованого) набору електронно-збуджених відносно Гартрі-Фоківського стану конфігурацій. Включаючи різну кількість конфігурацій з створеного набору, можливо отримати прогресивний набір наближених розв'язків до точних даних методу. Метод реалізовано в рамках теорії збурень Меллера-Плессета другого порядку (MP2) та різних рівнів теорії зв'язаних кластерів. Продемонстровано, що такі наближені розв'язки дають доволі точні значення енергетичних характеристик молекул, при цьому кількість конфігурацій у розрахунках може бути значно нижчою, ніж у розрахунках з використанням повного конфігураційного набору точного методу. Для ефективного розв'язку відповідних рівнянь теорії зв'язаних кластерів, реалізовано низку розрахункових алгоритмів з використанням багатокрокових методів першого порядку.

2. This thesis focuses on the study of the possibilities of L1-regularization application in the construction of "structure-activity" chemometric models and quantum chemical calculations. To perform the tasks of the thesis, an original set of programs has been developed that implement various statistical (chemometric) approaches to the construction of regression models and analysis of their prognostic properties. A set of quantum chemical programs has also been created, in which L1-regularization is used to construct wave functions of methods that take into account electronic correlation. In particular, in the thesis we consider application of L1-regularization to obtain linear empirical models for describing various physicochemical parameters of molecules. Based on the studied samples of molecules, it was shown that with the use of L1-regularization it is always possible to form a sequential (ordered) set of descriptors. By systematically adding descriptors from this set to linear regression models or artificial neural networks, it is possible to obtain regression models with successively increasing values of validation criteria. Due to the fact that after ranking of the descriptors set, the selected predictors can be used in different approaches to construct linear regression models, we conducted a corresponding study of the quality of these alternative models. It has been shown that the different methods can have better prognostic abilities

according to the criteria of external or internal validation. It is shown that with the use of artificial neural networks, based on the preliminary ordered by the method of L1-regularization descriptor set, high-quality predictions of the properties of matter can also be made. The obtained linear regression equations were also compared with alternative approaches that work with non-shrunked (non-optimized) descriptor sets. In the studied examples, we used L1-regularization to formulate compact one-, two- or three-parametric models that are able to satisfactorily describe the data set. According to the studied examples, the models obtained with pre-selection, using LARS-LASSO, turned out to be better than the results of PLS and PCR calculations. In the proposed PhD thesis some attention is paid to validation methods and quality of regression equations estimates. For this purpose, a model problem was used in which errors were introduced in both the dependent and independent variables. We considered the simplest case regression with one independent variable. It has been shown that random single sampling on the training and test sets is not informative. Therefore, in order to adequately estimate the quality of the regression equation, as well as to study the quality of the input data in general, it is necessary to create and study as many samplings into a training and test sample as possible. The known validation criteria proposed to date were also investigated. It is established that for data with substantial scatter the typical picture is the inverse (essentially nonlinear) dependence between external and internal validation criteria. Another problem that is closely related to the construction of statistical models is the construction of the classification function. For this purpose, the L1-regularized calculation of logistic regression was performed in this work. It is shown with the studied classification tasks that with the use of L1-regularized logistic regression it is possible to achieve classification results that are competitive with those obtained using other, more complex in the computational sense, methods. The use of a special L1-regularized algorithm made it possible to obtain fairly simple classification equations that are interpretable. Also, the obtained logistic regression equations are unambiguous and reproducible. It is shown that the L1-regularization method can be used in quantum chemistry. Using the L1-regularization procedure, it is possible to create an ordered (ranked) set of electronically excited configurations relative to the Gartree-Fock state. By including a different number of configurations from the created set, it is possible to obtain a progressive set of approximations to the exact calculations of the methods. The method is implemented in the framework of Meller-Plessett's theory of second-order perturbations (MP2) and different levels of the coupled clusters theory. It has been shown that such approximate solutions give fairly accurate values of the energy characteristics of molecules, and the number of configurations in the calculations can be much lower than in calculations using a complete configuration set of the exact method. A number of computational algorithms using first-order multi-step methods have been implemented to effectively solve the corresponding equations of the coupled clusters theory.

**Державний реєстраційний номер ДіР:**

**Пріоритетний напрям розвитку науки і техніки:**

**Стратегічний пріоритетний напрям інноваційної діяльності:**

**Підсумки дослідження:**

**Публікації:**

**Наукова (науково-технічна) продукція:**

**Соціально-економічна спрямованість:**

**Охоронні документи на ОПІВ:**

**Впровадження результатів дисертації:**

**Зв'язок з науковими темами:**

## **VI. Відомості про наукового керівника/керівників (консультанта)**

### **Власне Прізвище Ім'я По-батькові:**

1. Іванов Володимир Венедиктович
2. Ivanov Volodymyr V.

**Кваліфікація:** д. х. н., 02.00.04

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:**

**Код за ЄДРПОУ:**

**Місцезнаходження:**

**Форма власності:**

**Сфера управління:**

**Ідентифікатор ROR:** Не застосовується

## **VII. Відомості про офіційних опонентів та рецензентів**

### **Офіційні опоненти**

#### **Власне Прізвище Ім'я По-батькові:**

1. Кузьмін Віктор Євгенович
2. Kuz'min Victor E.

**Кваліфікація:** д. х. н., 02.00.03

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:**

**Код за ЄДРПОУ:**

**Місцезнаходження:**

**Форма власності:**

**Сфера управління:**

**Ідентифікатор ROR:** Не застосовується

#### **Власне Прізвище Ім'я По-батькові:**

1. Степаньян Степан Григорович
2. Stepanian Stepan Grygorovych

**Кваліфікація:** д. ф.-м. н., 01.04.14

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:**

**Код за ЄДРПОУ:**

**Місцезнаходження:**

**Форма власності:**

**Сфера управління:**

**Ідентифікатор ROR:** Не застосовується

**Рецензенти**

**Власне Прізвище Ім'я По-батькові:**

1. Пантелеймонов Антон Віталійович

2. Panteleimonov Anton Vitalijovych

**Кваліфікація:** к.х.н., 02.00.02

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:**

**Код за ЄДРПОУ:**

**Місцезнаходження:**

**Форма власності:**

**Сфера управління:**

**Ідентифікатор ROR:** Не застосовується

**Власне Прізвище Ім'я По-батькові:**

1. Черановський Владислав Олегович

2. Cheranovskii Vladyslav Olegovych

**Кваліфікація:** д.ф.-м.н., 01.04.07

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:**

**Код за ЄДРПОУ:**

**Місцезнаходження:**

**Форма власності:**

**Сфера управління:**

**Ідентифікатор ROR:** Не застосовується

## **VIII. Заключні відомості**

**Власне Прізвище Ім'я По-батькові  
голови ради**

Кириченко Олександр Васильович

**Власне Прізвище Ім'я По-батькові  
головуючого на засіданні**

Кириченко Олександр Васильович

**Відповідальний за підготовку  
облікових документів**

**Реєстратор**

**Керівник відділу УкрІНТЕІ, що є  
відповідальним за реєстрацію наукової  
діяльності**



Юрченко Т.А.