

MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN FEDERATION

**Federal State Autonomous Educational Institution of Higher Education
«National Research Lobachevsky State University of Nizhny Novgorod»**

Институт информационных технологий, математики и механики

УТВЕРЖДЕНО

решением Ученого совета ННГУ

протокол № 10 от 02.12.2024 г.

Working programme of the discipline

Machine learning

Higher education level

Master degree

Area of study / speciality

02.04.02 - Fundamental Informatics and Information Technology

Focus /specialization of the study programme

Artificial Intelligence and Data Analysis

Mode of study

full-time

Nizhny Novgorod

Year of commencement of studies 2025

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.О.05 Машинное обучение относится к обязательной части образовательной программы.

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства	
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	Для текущего контроля успеваемости	Для промежуточной аттестации
ОПК-2: Способен применять компьютерные/суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности	ОПК-2.1: Знает основные положения и концепции в области программирования, архитектуру языков программирования, теории коммуникации, знает основную терминологию, знаком с перечнем ПО, включенного в Единый Реестр Российских программ ОПК-2.2: Умеет анализировать типовые языки программирования, составлять программы ОПК-2.3: Имеет практический опыт решения задач анализа) интеграции различных типов программного обеспечения, анализа типов коммуникации	ОПК-2.1: ЗНАТЬ основные алгоритмы и методы машинного обучения основы языка Python или среды вычислений R ОПК-2.2: уметь работать с библиотекой Scikit-Learn или средой для статистических вычислений R ОПК-2.3: владеть практическим опытом работы с библиотекой Scikit-Learn или средой для статистических вычислений R	Задачи	Экзамен: Контрольные вопросы
ОПК-6: Способен к организации и ведению инновационно-исследовательской деятельности	ОПК-6.1: Обладает знаниями в области организации и ведения инновационно-исследовательской деятельности ОПК-6.2: Умеет осуществлять организацию и ведение инновационно-исследовательской деятельности ОПК-6.3: Имеет практический опыт организации и ведения	ОПК-6.1: ЗНАТЬ основные концептуальные и теоретические модели машинного обучения как инновационно-исследовательской деятельности ОПК-6.2: УМЕТЬ разрабатывать и анализировать концептуальные и	Тест	Экзамен: Контрольные вопросы

	инновационно-исследовательской деятельности	теоретические модели машинного обучения как инновационно-исследовательской деятельности ОПК-6.3: ИМЕТЬ НАВЫКИ оценивания на практике качество обучения инновационно-исследовательской моделей		
ПК-3: Способен эксплуатировать, разрабатывать и развивать системное и прикладное программное обеспечение, новые информационные технологии на основе анализа современного состояния науки и информационных технологий, и управлять такой эксплуатацией и разработкой в области профессиональной деятельности	ПК-3.1: Знает основы ИТ и имеет навыки анализа современного состояния науки и информационных технологий в области профессиональной деятельности ПК-3.2: Умеет применять навыки проектирования и разработки и развития ИТ-решений на основе анализа современного состояния науки и информационных технологий в области профессиональной деятельности ПК-3.3: Имеет практический опыт управления разработкой и развитием ИТ-решений на основе анализа современного состояния науки и информационных технологий в области профессиональной деятельности	ПК-3.1: Знать алгоритмы и методы машинного обучения ПК-3.2: Уметь использовать методы машинного обучения, оценивать качество методов ПК-3.3: Владеть современными программными средствами для решения задач машинного обучения: библиотекой Scikit-Learn, или средой языка Python или средой R для статистический вычислений; опытом реализации программных систем для решения практических задач с использованием методов машинного обучения	Тест	Экзамен: Контрольные вопросы

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная
Общая трудоемкость, з.е.	4
Часов по учебному плану	144
в том числе	
аудиторные занятия (контактная работа):	
- занятия лекционного типа	32
- занятия семинарского типа (практические занятия / лабораторные работы)	32

- КСР	2
самостоятельная работа	42
Промежуточная аттестация	36 Экзамен

3.2. Содержание дисциплины

(структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий)

Наименование разделов и тем дисциплины	Всего (часы)	в том числе			
		Контактная работа (работа во взаимодействии с преподавателем), часы из них			Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия семинарского типа (практические занятия/ лабора торные работы), часы	Всего	
	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0
Setting machine learning problems	34	10	10	20	14
Tutored training	34	10	10	20	14
Unsupervised learning	38	12	12	24	14
Аттестация	36				
КСР	2			2	
Итого	144	32	32	66	42

Contents of sections and topics of the discipline

Machine learning problems. Features. Supervised and unsupervised learning. Decision function (decision rule).

The generalizing ability of the decision function (the problem of the quality of learning). Classification, regression, clustering. Examples of practical tasks.

Probabilistic formulation of the machine learning problem. The principle of minimizing empirical risk. Bayesian decision theory. The principle of maximum posterior probability. Regression function. Bayesian classifier.method.

Experimental methods for assessing the quality of education. Separation of data into training and test samples. Cross-validation.

Least squares method. Maximum likelihood method. Linear regression model. System of normal equations.

The problem of overfitting when solving the regression problem. Methods to combat overfitting: reduction in the number of parameters, regularization (ridge regression), lasso. Nearest neighbors method. Risk estimation theorem. Naive Bayesian classifier. Linear Discriminant Analysis. Quadratic discriminant analysis. Logistic regression. Neural networks. Stochastic Gradient Descent. Backpropagation. gularization. Dropout. The concept of deep neural networks. Support Vector Machine. Kernel trick. Decision trees. CART method (classification and regression trees) for solving classification problems and regression reconstruction. Methods for handling missing values. Ensembles of decision rules (classifiers). Boosting. AdaBoost algorithm. Boosting and additive

models. Gradient boosting. Gradient Boosting Trees. Bagging. Random forests. Unsupervised learning. Clustering. K-means, k-medoids, Expectation-Maximization. DBSCAN Hierarchical clustering. Agglomerative and separating methods. Foundations of the Vapnik – Chervonenkis theory. Bernstein's lemma. A theorem on the uniform convergence of empirical risk to the expected risk in the case of a finite class of decision rules. Substantiation of the principle of minimizing empirical risk. Vapnik – Chervonenkis dimension. Sauer's lemma. A theorem on the uniform convergence of empirical risk to expected risk in the case of finite Vapnik – Chervonenkis dimension. The principle of structural risk minimization.

Практические занятия /лабораторные работы организуются, в том числе, в форме практической подготовки, которая предусматривает участие обучающихся в выполнении отдельных элементов работ, связанных с будущей профессиональной деятельностью.

На проведение практических занятий / лабораторных работ в форме практической подготовки отводится: очная форма обучения - 2 ч.

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся включает в себя подготовку к контрольным вопросам и заданиям для текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведенным в п. 5.

Electronic resources are recommended for preparing for practical exercises, independent work and intermediate certification <http://www.uic.unn.ru/~zny/ml/>

5. Assessment tools for ongoing monitoring of learning progress and interim certification in the discipline (module)

5.1 Model assignments required for assessment of learning outcomes during the ongoing monitoring of learning progress with the criteria for their assessment:

5.1.1 Model assignments (assessment tool - Tasks) to assess the development of the competency ОПК-2:

1. A training sample is given

x1 0 2 1 1 1 2 4 4 4 6

x2 4 4 3 5 4 1 0 2 1 1

y 0 0 0 0 0 1 1 1 1 1

Using linear discriminant analysis, construct a discriminant function for each class and write down the equation of the separating surface.

2. A training sample is given (see the table above). To construct discriminant functions using the method of quadratic discriminant analysis.

Assessment criteria (assessment tool — Tasks)

Grade	Assessment criteria
pass	The problem has been solved completely, or the main part of the problem has been solved, or the problem has been solved with shortcomings
fail	The problem has not been solved or the first stage of solving the problem has been completed

5.1.2 Model assignments (assessment tool - Test) to assess the development of the competency ПК-3:

1. Which of the following tasks is the task of learning without a teacher?

- a. regression reconstruction
- b. Classification
- c. clusterization

2. Which of the following tasks is the task of learning with a teacher?

- a. Classification
- b. clusterization
- c. restoration of the distribution function

5.1.3 Model assignments (assessment tool - Test) to assess the development of the competency ОПК-6:

1. Which of the following classification methods consists in constructing a separating hyperplane?

- a. logistic regression
- b. k nearest neighbors method
- c. random forest

2. Which of the following methods are insensitive to monotonic transformations of input parameters?

- a. the k nearest neighbors method
- b. the support vector
- method c. random forests

Assessment criteria (assessment tool — Test)

Grade	Assessment criteria
pass	51-100%
fail	0-50%

5.2. Description of scales for assessing learning outcomes in the discipline during interim certification

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатора достижения компетенций)	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	не зачтено		зачтено				
<u>Знания</u>	Отсутствие знаний теоретического материала. Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	Уровень знаний ниже минимальных требований. Имели место грубые ошибки	Минимально допустимый уровень знаний. Допущено много негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько несущественных ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Ошибок нет.	Уровень знаний в объеме, превышающем программу подготовки.
<u>Умения</u>	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки	Продemonстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продemonстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания в полном объеме, но некоторые с недочетами	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения. Решены все основные задачи с отдельным и несущественными недочетами, выполнены все задания в полном объеме	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
<u>Навыки</u>	Отсутствие базовых навыков. Невозможность оценить наличие навыков вследствие отказа обучающегося от	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для решения стандартных задач с некоторым	Продemonстрированы базовые навыки при решении стандартных задач с некоторым и	Продemonстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продemonстрированы навыки при решении нестандартных задач без ошибок и	Продemonстрирован творческий подход к решению нестандартных задач

	ответа		и недочетами	недочетами		недочетов	
--	--------	--	-----------------	------------	--	-----------	--

Scale of assessment for interim certification

Grade		Assessment criteria
pass	outstanding	All the competencies (parts of competencies) to be developed within the discipline have been developed at a level no lower than "outstanding", the knowledge and skills for the relevant competencies have been demonstrated at a level higher than the one set out in the programme.
	excellent	All the competencies (parts of competencies) to be developed within the discipline have been developed at a level no lower than "excellent",
	very good	All the competencies (parts of competencies) to be developed within the discipline have been developed at a level no lower than "very good",
	good	All the competencies (parts of competencies) to be developed within the discipline have been developed at a level no lower than "good",
	satisfactory	All the competencies (parts of competencies) to be developed within the discipline have been developed at a level no lower than "satisfactory", with at least one competency developed at the "satisfactory" level.
fail	unsatisfactory	At least one competency has been developed at the "unsatisfactory" level.
	poor	At least one competency has been developed at the "poor" level.

5.3 Model control assignments or other materials required to assess learning outcomes during the interim certification with the criteria for their assessment:

5.3.1 Model assignments (assessment tool - Control questions) to assess the development of the competency ОПК-2

1. Machine learning problems. Features. Supervised and unsupervised learning. Decision function (decision rule). The generalizing ability of the decision function (the problem of the quality of learning). Classification, regression, clustering.
2. Probabilistic formulation of the machine learning problem. The principle of minimizing empirical risk. Bayesian decision theory. The principle of maximum posterior probability. Regression function. Bayesian classifier.method.
3. Experimental methods for assessing the quality of education. Separation of data into training and test samples. Cross-validation.
4. Least squares method. Maximum likelihood method. Linear regression model. System of normal equations.
5. The problem of overfitting when solving the regression problem. Methods to combat overfitting: reduction in the number of parameters, regularization (ridge regression), lasso.

5.3.2 Model assignments (assessment tool - Control questions) to assess the development of the competency ПК-3

1. Logistic regression. Neural networks. Stochastic Gradient Descent. Backpropagation. Regularization. Dropout. The concept of deep neural networks.
2. Support Vector Machine. Kernel trick.
3. Decision trees. CART method (classification and regression trees) for solving classification problems and regression reconstruction. Methods for handling missing values.
4. Ensembles of decision rules (classifiers). Boosting. AdaBoost algorithm. Boosting and additive models. Gradient boosting. Gradient Boosting Trees. Bagging. Random forests.
5. Unsupervised learning. Clustering. K-means, k-medoids, Expectation Maximization. DBSCAN
6. Hierarchical clustering. Agglomerative and separating methods.
7. Foundations of the Vapnik – Chervonenkis theory. Bernstein's lemma. A theorem on the uniform convergence of empirical risk to the expected risk in the case of a finite class of decision rules. Substantiation of the principle of minimizing empirical risk. Vapnik – Chervonenkis dimension. Sauer's lemma. A theorem on the uniform convergence of empirical risk to expected risk in the case of finite Vapnik – Chervonenkis dimension. The principle of structural risk minimization.

5.3.3 Model assignments (assessment tool - Control questions) to assess the development of the competency ОПК-6

1. Nearest neighbors method. Risk estimation theorem.
2. Naive Bayesian classifier.
3. Linear Discriminant Analysis. Quadratic discriminant analysis.

Assessment criteria (assessment tool — Control questions)

Grade	Assessment criteria
outstanding	fluency in basic and additional material without mistakes and errors
excellent	fluent knowledge of the basic material with minor errors and mistakes
very good	sufficient knowledge of the basic material with minor errors
good	mastery of the basic material with a number of noticeable errors
satisfactory	knowledge of the minimum material required for a given subject, with a number of errors
unsatisfactory	knowledge of the material is insufficient, additional preparation is required. The work during the semester can be assessed as unsatisfactory
poor	lack of knowledge of the material. The work during the semester was assessed as "poor"

Grade	Assessment criteria

6. Учебно-методическое и информационное обеспечение дисциплины (модуля)

Основная литература:

1. Sebastian Raschka. Python Machine Learning : Machine Learning and Deep Learning with Python, Scikit-learn, and TensorFlow 2, 3rd Edition. - Packt Publishing, 2019. - 1 online resource. - ISBN 9781789958294. - ISBN 9781789955750. - Текст : электронный., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=854412&idb=0>.

Дополнительная литература:

1. Jesus Salcedo. Machine Learning for Data Mining : Improve Your Data Mining Capabilities with Advanced Predictive Modeling. - Packt Publishing, 2019. - 1 online resource. - ISBN 9781838821555. - ISBN 9781838828974. - Текст : электронный., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=854274&idb=0>.

2. Amita Kapoor. Hands-On Artificial Intelligence for IoT : Expert Machine Learning and Deep Learning Techniques for Developing Smarter IoT Systems. - Packt Publishing, 2019. - 1 online resource. - ISBN 9781788832762. - ISBN 9781788836067. - Текст : электронный., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=854227&idb=0>.

Программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

1. The R Project for Statistical Computing <https://www.r-project>
2. Welcome to Python.org <https://www.python.org/>
3. scikit-learn: machine learning in Python scikit-learn.org/

7. Материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

Программа составлена в соответствии с требованиями ОС ННГУ по направлению подготовки/специальности 02.04.02 - Fundamental Informatics and Information Technology.

Авторы: Золотых Николай Юрьевич, доктор физико-математических наук, доцент.

Заведующий кафедрой: Золотых Николай Юрьевич, доктор физико-математических наук.

Программа одобрена на заседании методической комиссии от 02.12.2024, протокол № 5.