

**Федеральное государственное автономное образовательное учреждение
высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"**

Факультет компьютерных наук
Кафедра технологий моделирования сложных систем

**Рабочая программа дисциплины
«Геометрические методы предсказательного моделирования»**

для образовательной программы «Науки о данных»
направления подготовки 01.04.02 Прикладная математика и информатика
магистратура

Разработчик программы

Янович Ю.А., к.ф.-м.н., ст.преподаватель кафедры технологий моделирования сложных систем

Одобрена на заседании кафедры технологий моделирования сложных систем

«__»_____ 201_ г.

Зав. кафедрой технологий моделирования сложных систем

А.Н.Соболевский

Утверждена Академическим советом образовательной программы

«__»_____ 201_ г., № протокола _____

Академический руководитель образовательной программы

«Науки о данных» С.О.Кузнецов

Москва, 2018

Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения подразделения разработчика программы

1 Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает требования к образовательным результатам и результатам обучения студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих дисциплину «Введение в технологию блокчейн», учебных ассистентов и студентов направления подготовки 01.04.02 Прикладная математика и информатика, обучающихся по образовательной программе «Науки о данных».

Программа учебной дисциплины разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ по направлению подготовки 01.04.02 Прикладная математика и информатика;
- Образовательной программой «Науки о данных».
- Объединенным учебным планом университета по образовательной «Науки о данных», утвержденным в 2017г.

2 Цели освоения дисциплины

Целью освоения курса является изучение системы современных методов построения эффективных математических моделей поведения сложных многокомпонентных систем. Сутью методологии предсказательного моделирования является построение «облегченных» математических моделей поведения сложных многокомпонентных систем, позволяющих сочетать простоту вычислений по модели с достаточной точностью и надежностью. Этот подход является альтернативой традиционному подходу в математическом моделировании, когда модель строится «из первых принципов» как адекватное математическое описание собственной динамики системы. Предсказательное моделирование опирается на такие области математики, как теория аппроксимации, обучение машин, математическая статистика, теория информации, теория игр.

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

- знать основные понятия и методы математической статистики, относящиеся к анализу многомерных данных;
- понимать природу изучаемых методов и их место в общей системе теоретико-вероятностного и статистического знания;
- уметь применять изученные статистические методы для анализа многомерных данных.

В результате освоения дисциплины студент осваивает следующие компетенции:

Компетенция	Код по ФГОС/ НИУ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
Способность использовать в профессиональной деятельности зна-	ИК-М7.1 пми	РБ	Студент дает определение понятий, применяет изученные методы при решении задач	Стандартные (лекционно-семинарские). Самостоятельные	итоговый экзамен

Компетенция	Код по ФГОС/ НИУ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
ния в области предсказательного моделирования				внеаудиторные занятия.	
Способность строить и решать статистические модели	ИК-М7.2 пми	СД	Студент дает определение понятий, применяет изученные методы при решении задач	Стандартные (лекционно-семинарские). Самостоятельные внеаудиторные занятия.	Домашнее задание, итоговый экзамен
Способность понимать и применять в исследовательской и прикладной деятельности аппарат теории вероятностей и математической статистики	ИК-М7.3 пми	СД	Студент дает определение понятий, применяет изученные методы при решении задач	Стандартные (лекционно-семинарские). Самостоятельные внеаудиторные занятия.	Домашнее задание, итоговый экзамен

4 Место дисциплины в структуре образовательной программы

Для специализации «Технологии моделирования сложных систем» настоящая дисциплина является базовой. Изучение курса «Геометрические методы предсказательного моделирования» требует предварительных знаний в объеме первых курсов стандартной бакалаврской программы по этой (010500.62) или смежной тематике.

5 Тематический план учебной дисциплины

№	Название раздела	Всего часов	Аудиторные часы		Самостоятельная работа
			Лекции	Практич. занятия	
1	Элементы теории статистических решений	28	4	6	18
2	Элементы теоретической математической статистики	30	4	6	20
3	Основные модели многомерных данных	28	4	6	18

4	Обзор основных задач анализа многомерных данных	30	4	6	20
5	Линейный регрессионный анализ	28	4	6	18
	Итого	144	30	20	94

6 Формы контроля знаний студентов

Тип контроля	Форма контроля	1	2	Параметры
Текущий (неделя)	Домашнее задание	4		
	Проект		1	
Итоговый	Экзамен		*	устный

Порядок формирования оценок по дисциплине

Преподаватель оценивает правильность выполнения домашних заданий студентов. Результирующая оценка по 10-ти балльной шкале за самостоятельную работу определяется перед итоговым контролем – $O_{дз}$ как среднее значение оценок за отдельные домашние задания. Задания, присланные с опозданием менее 30 дней и не менее 7 дней до экзамена, оцениваются с дополнительным множителем 0,5.

В середине курса студентам предлагаются темы исследовательских и программистских проектов. В конце курса производится их публичная защита, результирующая оценка по 10-ти балльной шкале за которые определяются как $O_{пр}$.

В диплом выставляется результирующая оценка по учебной дисциплине, которая формируется по следующей формуле, где $O_{экзамен}$ – оценка за ответ на устном экзамене:

$$O_{результ} = 0,3 \cdot O_{дз} + 0,3 \cdot O_{пр} + 0,4 \cdot O_{экзамен}.$$

Способ округления оценок: арифметический.

7 Содержание дисциплины

Тема 1. Элементы теории статистических решений.

Элементы теории статистических решений как теоретической основы предсказательного моделирования. Основные понятия: пространства состояний, наблюдений, решений, действий; решающие функции; функции потерь и функции риска; наилучшие и допустимые решающие функции; полные классы; множества Парето; минимаксный и байесовский риски; байесовские решающие функции.

Тема 2. Элементы теоретической математической статистики.

Проверка статистических гипотез (простые и сложные гипотезы; критерии; уровень значимости и функции мощности; наиболее мощные критерии и теорема Неймана-Пирсона; критерии для проверки гипотез о параметрах нормального распределения; непараметрические критерии).

Теория статистического оценивания (оценки и их меры качества; несмещенные оценки и

оценки с наименьшей дисперсией; неравенство Рао-Крамера; оценки параметров нормального распределения; оценки плотности и функции распределения).

Исключение резко выделяющихся наблюдений (различные критерии).

Тема 3. Основные модели многомерных данных.

Основные модели многомерных данных. Частные и условные распределения. Многомерное нормальное распределение. Распределение линейных комбинаций и квадратичных форм от компонент многомерного нормального вектора. Зависимости между компонентами многомерного нормального вектора. Частные и множественные коэффициенты корреляции. Зависимости между компонентами многомерного нормального вектора. Частные и множественные коэффициенты корреляции. Теорема о нормальной корреляции.

Тема 4. Обзор основных задач анализа многомерных данных.

Обзор основных задач анализа многомерных данных (линейный и нелинейный регрессионный анализ, факторный анализ, корреляционный анализ, дисперсионный анализ, снижение размерности).

Тема 5. Линейный регрессионный анализ.

Линейный регрессионный анализ. Обычный и обобщенный методы наименьших квадратов. Геометрическая интерпретация. Регрессионный анализ для мультиколлинеарных данных. Оценка ошибок линейной модели. Проверка гипотез о параметрах линейной модели. Доверительные интервалы.

Вычислительные аспекты линейного регрессионного анализа. Регуляризация. Пошаговые методы.

8 Учебно-методическое и информационное обеспечение дисциплины

1. Закс Ш. Теория статистических выводов. М.: Мир, 1975.
2. Крамер Г. Математические методы статистики. М.: Мир, 1975.
3. Андерсон Т. Введение в многомерный статистический анализ. М.: Физматгиз, 1963.
4. Себер Дж. Линейный регрессионный анализ. М.: Мир, 1980.
5. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Основы моделирования и первичная обработка данных. М.: Финансы и статистика, 1983.
6. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Исследование зависимостей. М.: Финансы и статистика, 1985.
7. Айвазян С.А., Бухштабер В.М., Енюков С.А., Мешалкин Л.Д. Прикладная статистика. Классификация и снижение размерности. М.: Финансы и статистика, 1989.
8. Бокс Дж., Дженкинс Г. Анализ временных рядов, прогноз и управление. М: Мир, 1974.
9. Гнеденко Б. В., Коваленко И. Н. Введение в теорию массового обслуживания. М.: Наука, 1966.
10. Оуэн Г. Теория игр. М.: Мир, 1971.