



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

Факультет Компьютерных наук
Департамент больших данных и информационного поиска

Рабочая программа дисциплины
Современные методы анализа данных: Глубинное обучение

для образовательной программы «Финансовые технологии и анализ данных»
направления подготовки 01.04.02 «Прикладная математика и информатика»
уровень магистр

Разработчик(и) программы
Осокин А.А., к.ф.-м.н., aosokin@hse.ru

Одобрена на заседании департамента больших данных и информационного поиска
«__»_____ 2017 г.

Руководитель департамента
Подольский В. В. _____

Утверждена Академическим советом образовательной программы
«__»_____ 2017 г., № протокола _____

Академический руководитель образовательной программы
Масютин А.А. _____

Москва, 2017

*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 «Современные методы анализа данных: Глубинное обучение», учебных ассистентов и студентов направления подготовки 01.04.02 «Прикладная математика и информатика», обучающихся по образовательной программе «Финансовые технологии и анализ данных». Программа учебной дисциплины разработана в соответствии с:

- образовательным стандартом Федерального государственного автономного образовательного учреждения высшего образования Национальный исследовательский университет «Высшая школа экономики», уровень подготовки: магистр, утвержденным ученым советом НИУ ВШЭ 22.12.2017 протокол №13;
- рабочим учебным планом университета по направлению 01.04.02 «Прикладная математика и информатика» подготовки магистра для магистерской программы «Финансовые технологии и анализ данных», утвержденным 02.02.2018 г.

2 Цели освоения дисциплины

Целями освоения дисциплины «Современные методы анализа данных: Глубинное обучение» являются:

- Ознакомление студентов с основными принципами применения методов, основанных на глубоких нейросетях, для решения задач машинного обучения.
- Формирование у студентов практических навыков применения и обучения глубоких нейросетей в прикладных задачах анализа данных из таких областей как компьютерное зрение, обработка текстов, и др.

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

Уровни формирования компетенций:

РБ — ресурсная база, в основном теоретические и предметные основы (знания, умения);

СД – способы деятельности, составляющие практическое ядро данной компетенции;

МЦ – мотивационно-ценностная составляющая, отражает степень осознания ценности компетенции человеком и готовность ее использовать

В результате освоения дисциплины студент осваивает компетенции:

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
	УК-1	РБ	Способен рефлексировать (оценивать и перерабатывать) освоенные научные методы и способы деятельности.	Выполнение домашних заданий и проектной работы	Домашние задания и проект, экзамен



Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
	УК-2	РБ	Способен создавать новые теории, изобретать новые способы и инструменты профессиональной деятельности.	Выполнение проектной работы	Проект
	УК-3	РБ	Способен к самостоятельному освоению новых методов исследований, изменению научного и производственного профиля своей деятельности.	Выполнение проектной работы	Проект
	ОПК-1	СД	Способен применять системный подход при постановке задач и выборе подходов к решению, а также для учёта противоречивых целей, потребностей и требований.	Выполнение проектной работы	Проектная работа, экзамен
	ОПК-3	СД	Способен обоснованно выбирать и применять в профессиональной деятельности современные компьютерные технологии в соответствии со спецификой решаемых задач, включая операционные системы, сетевые технологии, языки программирования, языки манипулирования данными, электронные библиотеки, пакеты прикладных программ.	Выполнение проектной работы	Проектная работа
	ОПК-5	СД	Способен строить математические модели и использовать их при решении прикладных задач в соответствии с направлением подготовки и специализацией.	Выполнение домашних заданий и проектной работы	Домашние задания, проект
	ПК-7	СД	Способен осуществлять целенаправленный многокритериальный поиск информации о новейших	Выполнение домашних заданий и проектной работы	Домашние задания, проект



Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
			научных и технологических достижениях в сети Интернет и в других источниках.		
	ПК-10	СД	Способен реализовывать модели и алгоритмы прикладной математике в виде компьютерных программ.	Выполнение домашних заданий и проектной работы	Домашние задания, проект
	ПК-11	СД	Способен оценивать корректность и воспроизводимость применения методов прикладной математики и информатики	Выполнение домашних заданий и проектной работы	Домашние задания, проект

4 Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к циклу дисциплин по машинному обучению и анализу данных.

Изучение данной дисциплины базируется на следующих дисциплинах:

- Математический анализ
- Линейная алгебра и геометрия
- Теория вероятностей
- Алгоритмы и структуры данных
- Машинное обучение 1
- Машинное обучение 2

5 Тематический план учебной дисциплины

№	Название раздела	Всего часов	Аудиторные часы		Самостоятельная работа
			Лекции	Семинары	
1	Введение в курс	12	2	2	8
2	Блок тем 1: Основные компоненты глубинного обучения: модели и методы обучения	36	6	6	24
3	Блок тем 2: Области применимости глубинного обучения: компьютерное зрение, обработка естественных языков, вероятностные модели	66	12	12	42
	ИТОГО	114	20	20	74



6 Формы контроля знаний студентов

Тип контроля	Форма контроля	1 год	Параметры
		3	
Текущий	Домашнее задание	*	Практическая работа: вывод формул, программная реализация модели, проведение экспериментов, анализ результатов эксперимента, написание полного отчета
	Проект	*	Решение практической задачи на основе материалов курса
Итоговый	Экзамен	*	Устно-письменный, состоящий в ответах на вопросы в письменной и устной формах

7 Критерии оценки знаний, навыков

В курсе предусмотрено несколько форм контроля знания:

- Домашние задания по материалам лекций и семинаров. Задания могут состоять в доведении до конца решений задач с семинаров, решении теоретических задач, решении практических задач.
- Проект, связанный с реализацией метода глубинного обучения.
- Устно-письменный экзамен.

Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

8 Содержание дисциплины

1. Введение в глубинное обучение

Нейронные сети для задачи классификации, оптимизация методом стохастического градиента, примеры задач.

Блок тем 1.



2. Механика нейросетей и алгоритм обратного распространения ошибок

Обратное распространение ошибки как основной способ обучения нейросетей, дифференцирование основных операций (полносвязные слои, свёртка), автоматическое дифференцирование сложных моделей, подходы к реализации нейросетевых библиотек (статический и динамический графы вычислений)

3. Основные виды нейросетей

Повышение эффективности использования параметров, операция свёрки для изображений, основные принципы построения свёрточных сетей, рекуррентные сети для обработки последовательностей, основные виды рекуррентных блоков и способов построения рекуррентных моделей.

4. Обучение и регуляризация нейросетей

Методы регуляризации нейросетей (L2, dropout, batchnorm, data augmentation и др.), методы оптимизации (SGD и его улучшения, подбор длины шага и масштаба каждого из параметров), архитектурные элементы, улучшающие обучение нейросетей (gating, skip connections).

Блок тем 2.

5. Глубинное обучение для задач компьютерного зрения

Примеры построения сложных архитектур для задач компьютерного зрения: поиск объектов на изображении (object detection), сегментация изображений (image segmentation), обучение представлений и поиск изображений (representation learning and image retrieval), и др.

6. Глубинное обучение для задач обработки текстов

Представления слов (word embeddings), модели для предсказания последовательностей (sequence-to-sequence), задачи машинного перевода и генерации подписи к изображениям.

7. Глубинное обучение для построения вероятностных моделей

Прямые вероятностные модели (NADE, PixelCNN и др.), вероятностные модели со скрытыми переменными (вариационный автокодировщик, VAE).



8. Непрямые способы использования нейросетей

Адаптация нейросетей к новым наборам данных (domain adaptation), противоборствующее обучение (adversarial trainings), построение примеров, на которых нейросети ошибаются (adversarial examples).

9. Встраивание алгоритмов в слои нейросетей

Алгоритм вывода в функции потерь на примере структурного метода опорных векторов, дифференцируемый вывод на примере гауссовского марковского поля, итерационные алгоритмы как вычислительные графы на примере алгоритма передачи сообщений.

10. Недифференцируемые модели и глубинное обучение с подкреплением

Введение в обучение с подкреплением и основные алгоритмы (Q-learning и policy gradients), глубинное обучение с подкреплением, байесовские нейросети.

11. Приглашенная лекция представителя индустрии с разбором реального проекта, использующего глубинное обучение

9 Образовательные технологии

Чтение лекций и проведение семинаров. Семинары посвящены практическому применению материала, разобранным на лекциях. Задания с семинара, не выполненные на самом семинаре, выносятся на домашнее задание.

Методические указания студентам. Домашние задания выдаются каждую неделю, поэтому рекомендуется их выполнять на регулярной основе, чтобы предотвратить эффект снежного кома нагрузки. Также некоторые домашние задания могут подразумевать значительные временные затраты на вычислительные эксперименты и неожиданные трудности, поэтому рекомендуется не откладывать начало выполнения домашних заданий на последний момент.

10 Оценочные средства для текущего контроля и аттестации студента

Примеры домашних задания

1. Разработать, запрограммировать, применить нейросетевой алгоритм распознавания рукописных символов для базы данных MNIST.
2. Разработать, запрограммировать, применить нейросетевой алгоритм синтеза рукописных символов на основе подхода GAN. Провести оценку качества работы метода.



Примеры экзаменационных вопросов

1. Нейронные сети для задачи классификации, оптимизация методом стохастического градиента, примеры задач.
2. Механика нейросетей и алгоритм обратного распространения ошибок.
3. Принципы построения архитектур свёрточных сетей.
4. Принципы построения архитектур рекуррентных сетей.
5. Методы обучения и регуляризации нейросетей
6. Методы решения задачи поиска объектов на изображении.
7. Методы решения задачи машинного перевода
8. Методы построения глубинных вероятностных моделей
9. Методы противоборствующего обучения
10. Методы адаптации нейросетей к новым наборам данных
11. Методы построения слоев нейросетей на основе существующих алгоритмов
12. Методы глубинного обучения для задачи обучения с подкреплением

11 Порядок формирования оценок по дисциплине

В курсе предусмотрено 10 домашних заданий, выдаваемых еженедельно по итогам семинара, 1 проект, итоговый экзамен в конце курса.

Каждое задание и проект оцениваются по 10-балльной шкале (по заданиям допускается дробная оценка). Оценки, полученные за домашние задания, складываются и делятся на максимальный балл равный $10 \cdot \text{число заданий}$.

Накопленная оценка вычисляется по формуле

$$O_{\text{накопл}} = \text{округление}(0.6 O_{\text{задания}} + 0.4 O_{\text{проект}})$$

Итоговая оценка вычисляется по формуле

$$O_{\text{итог}} = 0.7 O_{\text{накопл}} + 0.3 O_{\text{экзамен}}$$

Способ округления результирующей оценки по учебной дисциплине: арифметический

Студенту, получившему отличную накопленную оценку, данная оценка может быть выставлена в качестве итоговой по усмотрению преподавателя.

12 Учебно-методическое и информационное обеспечение дисциплины

12.1 Базовый учебник

Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning, MIT Press, 2016.

Электронная версия учебника доступна онлайн: <http://www.deeplearningbook.org/>



12.2 Основная литература

- Документация библиотеки pytorch: <http://pytorch.org/docs/master/>
- Обучающие материалы библиотеки pytorch: <http://pytorch.org/tutorials/>

12.3 Дополнительная литература

Общие материалы

- Материалы курса Neural Networks из Шербрукского университета (лектор Hugo Larochelle), доступны все материалы и видеозаписи лекций:
http://info.usherbrooke.ca/hlarochelle/neural_networks/content.html

Глубинное обучение в компьютерном зрении

- Материалы курса CS231n: Convolutional Neural Networks for Visual Recognition из Стэнфордского университета (лекторы Fei-Fei Li, Justin Johnson, Serena Yeung), все материалы курса доступны на <http://cs231n.github.io/>
- Материалы курса CSC2523: Deep Learning in Computer Vision из университета Торонто (координатор курса Sanja Fidler), все материалы курса доступны на <http://www.cs.utoronto.ca/~fidler/teaching/2015/CSC2523.html>

Глубинное обучение для обработки естественного языка

- Материалы курса CS224n: Natural Language Processing with Deep Learning из Стэнфордского университета (лекторы Chris Manning, Richard Socher), все материалы курса доступны на <http://web.stanford.edu/class/cs224n/syllabus.html>

Глубинное обучение с подкреплением

- Материалы курса Deep Reinforcement Learning and Control университета Карнеги-Меллон (лекторы Katerina Fragkiadaki, Ruslan Satakhutdinov), все материалы курса доступны на <https://katefvision.github.io/>
- Материалы курса CS 294: Deep Reinforcement Learning из Калифорнийского университета в Беркли (лектор Sergey Levine), материалы курса доступны на <http://rl.berkeley.edu/deeprlcourse/>