**Программа учебной дисциплины майнор
«Современные методы машинного обучения»**

Утверждена

Академическим советом ООП

Протокол № от «\_\_»\_\_\_\_\_20\_\_ г.

|  |  |
| --- | --- |
| Автор  | Зимовнов Андрей Вадимович |
| Число кредитов  |  |
| Контактная работа (час.)  | 40 |
| Самостоятельная работа (час.)  | 40 |
| Курс  |  |
| Формат изучения дисциплины | Без использования онлайн курса |

1. **ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ**

Целью освоения данной дисциплины является знакомство с современными методами машинного обучения, активно использующимися сегодня на производстве.

В результате освоения дисциплины студент должен:

* знать устройство современных моделей машинного обучения и их особенности;
* владеть инструментами для построения этих моделей;
* уметь применять эти модели на практике.

Изучение данной дисциплины базируется на следующих дисциплинах:

- математика в объеме средней школы;

- введение в машинное обучение.

# Содержание УЧЕБНОЙ ДИСЦИПЛИНЫ

1. Метод опорных векторов. Ядра и ядровой трюк (kernel trick). Ядровой метод главных компонент. Области применимости ядровых методов: объекты без признаков, ядра в метрических и линейных методах.
2. Градиентный спуск (напоминание), обсуждение его особенностей и проблем. Стохастический градиентный спуск. Ускорение сходимости стохастических методов: Moment и AdaGrad. Метод Ньютона.
3. Метод AdaBoost. Градиентный бустинг над решающими деревьями. Сравнение бустинга со случайными лесами.
4. Определение нейронной сети. Обучение нейронных сетей, метод обратного распространения ошибки.
5. Введение в TensorFlow (пакет для обучения нейросетей).
6. Сверточные нейронные сети, современные архитектуры.
7. Рекуррентные нейронные сети для текстов.
8. Автокодировщики и другие применения нейросетей.
9. Проверка статистических гипотез, основные понятия: уровень значимости, достигаемый уровень значимости (p-value), ошибки I и II рода.
10. Временной ряд. Тренд, сезонность. Автокорреляционная функция. Коррелограмма и её интерпретация. Модели класса ARIMA.

# ОЦЕНИВАНИЕ

Курс подразумевает следующие способы контроля знаний:

* Домашние работы
* Проверочные на семинарах (письменные)
* Коллоквиум (по всему материалу 1-го модуля, письменный)
* Экзамен (в конце курса, письменный).

Результирующая оценка по дисциплине рассчитывается по формуле:

$$O\_{итог}=0.7 O\_{накопл}+0.3 O\_{экз}$$

Накопленная и итоговая оценки округляются арифметически.

Накопленная оценка рассчитывается по формуле:

$$O\_{накопл}=0.1 O\_{самост}+0.7 O\_{дз}+0.2 O\_{коллоквиум}$$

Оценка за домашние задания рассчитывается как среднее значение оценок за все выданные домашние задания. Оценка за самостоятельную работу рассчитывается как среднее значение оценок за все проверочные работы, проведённые на семинарских занятиях.

Студенту, получившему отличную накопленную оценку, данная оценка может быть выставлена в качестве итоговой на усмотрение семинариста и лектора.

Блокирующие элементы оценки отсутствуют.

# ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

Примеры экзаменационных вопросов:

* Бустинг для задачи регрессии, связь с градиентом функции потерь.
* Градиентный бустинг для произвольной функции потерь.
* Построение градиентного спуска с помощью ряда Тейлора.
* Инициализация весов в нейронных сетях. Можно ли нулями и почему. Из каких соображений лучше выбирать масштаб случайных весов.
* Устройство простой рекуррентной сети (две формулы с объяснением мотивации).
1. **РЕСУРСЫ**
	1. **Основная литература**
* Ian Goodfellow, Yoshua Bengio. Deep Learning, 2016.
(<http://www.deeplearningbook.org>)
* James, Witten, Hastie, Tibshirani. An Introduction to Statistical Learning, 2013. ([http://www-bcf.usc.edu/~gareth/ISL/ISLR Sixth Printing.pdf](http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Sixth%20Printing.pdf))
* Bishop C.M. Pattern Recognition and Machine Learning, 2006. (<https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>)
	1. **Дополнительная литература**
* Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms. Cambridge University Press, 2014.
* Diez, D.M, Barr, C.D., Cetinkaya-Rundel, M., Dorazio, L. Advanced High School Statistics. — OpenIntro, 2015.
	1. **Программное обеспечение**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Наименование** | **Условия доступа** |
| 1. |  Microsoft Windows 7 Professional RUSMicrosoft Windows 10Microsoft Windows 8.1 Professional RUS | *Из внутренней сети университета (договор)* |
| 2. | Microsoft Office Professional Plus 2010 | *Из внутренней сети университета (договор)* |

* 1. **Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Наименование** | **Условия доступа** |
|  | ***Профессиональные базы данных, информационно-справочные системы*** |
| 1. | Консультант Плюс | *Из внутренней сети университета (договор)* |
| 2. | Электронно-библиотечная система Юрайт  | URL: https://biblio-online.ru/ |
|  | ***Интернет-ресурсы (электронные образовательные ресурсы)*** |
| 1. | Открытое образование  | URL: https://openedu.ru/ |

* 1. **Материально-техническое обеспечение дисциплины**

Учебные аудитории для лекционных занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

* ПЭВМ с доступом в Интернет (операционная система, офисные программы,  антивирусные программы);
* мультимедийный проектор с дистанционным управлением.

Учебные аудитории для семинарских и самостоятельных занятий по дисциплине не требуют специального технического оснащения.