**Программа учебной дисциплины \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Утверждена

Академическим советом ООП

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

|  |  |
| --- | --- |
| Автор  | Соколов Евгений Андреевич |
| Число кредитов  | 5 |
| Контактная работа (час.)  | 80 |
| Самостоятельная работа (час.)  | 110 |
| Курс  | Машинное обучение 2 |
| Формат изучения дисциплины | Без использования онлайн курса |

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

Целями освоения дисциплины « Машинное обучение 2» являются:

* + Ознакомление студентов с современными методами машинного обучения — в частности, с методами обучения без учителя, генеративными моделями, активным обучением и т.д.
	+ Формирование у студентов практических навыков работы с данными и решения прикладных задач анализа данных.

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

Для освоения учебной дисциплины студенты должны владеть знаниями и компетенциями следующих дисциплин:

* Математический анализ
* Линейная алгебра и геометрия
* Теория вероятностей
* Математическая статистика
* Алгоритмы и структуры данных
* Машинное обучение 1

Основные положения дисциплины должны быть использованы в дальнейшем при изучении дисциплин:

• Машинное обучение на больших данных

• Байесовские методы машинного обучения

• Глубинное обучение

# Содержание УЧЕБНОЙ ДИСЦИПЛИНЫ

1. Обучение с учителем
2. Обучение без учителя
3. Метрические методы
4. Метод k ближайших соседей в классификации и регрессии. Метрики на сложных объектах. Метрики на текстах на основе представлений слов. Методы поиска ближайших соседей. Locality-sensitive hashing. Обучение метрик, learning to hash.
5. Ранжирование
6. Задача адаптации признакового пространства (domain adaptation), transfer learning
7. Оптимизация гиперпараметров, гауссовские процессы в оптимизации
8. Теория статистического обучения

# ОЦЕНИВАНИЕ

В рамках курса предусмотрены самостоятельные работы на занятиях, теоретические домашние задания, практические домашние задания, письменная контрольная работа и письменный экзамен.

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=0.2 O\_самост + 0.4 O\_практ + 0.2 О\_теор + 0.2 O\_контрольные

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. В конце семестра разрешается переписать все самостоятельные работы, пропущенные по уважительной причине.

# ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

Примеры практических заданий можно найти по ссылке http://wiki.cs.hse.ru/Машинное\_обучение\_2/2017\_2018#.D0.9F.D1.80.D0.B0.D0.BA.D1.82.D0.B8.D1.87.D0.B5.D1.81.D0.BA.D0.B8.D0.B5\_.D0.B7.D0.B0.D0.B4.D0.B0.D0.BD.D0.B8.D1.8F

Примеры вопросов к экзамену:

1. Теорема о представлении (Representer theorem) с доказательством.
2. Аппроксимация спрямляющего пространства: метод Нистрома и метод случайных признаков Фурье.
3. Скрытые переменные, полное и неполное правдоподобие на примере смеси распределений. EM-алгоритм (описание шагов).
4. Дивергенция Кульбака-Лейблера, её неотрицательность. Вывод E- и M-шагов через разложение логарифма неполного правдоподобия на нижнюю оценку и KL-дивергенцию.
5. Ядровой метод главных компонент.
6. Двойственная задача SVM (с выводом). Какие объекты называются периферийными, опорными граничными, опорными нарушителями?
7. Лапласиан графа, его свойства. Связь кратности нулевого собственного значения с числом компонент связности (с доказательством). Алгоритм спектральной кластеризации.
8. Методы частичного обучения: self-training, моделирование смесью гауссиан, трансдуктивный SVM, графовый подход на основе лапласианов.
9. Несбалансированная классификация: undersampling, oversampling, SMOTE.
10. Одноклассовая классификация: непараметрические методы (восстановление плотности), одноклассовый SVM.
11. User-based и item-based рекомендательные системы.
12. Модели со скрытыми переменными (latent factor model, LFM) для построения рекомендаций. Обучение LFM: стохастический градиентный спуск, ALS, HALS.
13. Неявная информация в рекомендательных системах, implicit ALS.
14. Факторизационные машины. FFM. Способы обучения моделей.
15. Метрики качества рекомендаций.
16. Оценки максимального правдоподобия на параметры многомерного нормального распределения.
17. All-subsequences-kernel, алгоритм его вычисления за О(|s||t|).
18. Вероятностная постановка задачи классификации, функционал среднего риска. Оптимальное байесовское решающее правило, наивный байесовский классификатор.
19. **РЕСУРСЫ**
	1. **Основная литература**
20. 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))
21. 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**
22. Boyd, Vandenberghe. Convex Optimization (http://stanford.edu/~boyd/cvxbook/bv\_cvxbook.pdf)
	1. **Программное обеспечение**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Наименование** | **Условия доступа** |
| 1. | Anaconda | *Свободно распространяемое ПО* |

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

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Наименование** | **Условия доступа** |
|  | ***Интернет-ресурсы (электронные образовательные ресурсы)*** |
| 1. | Открытое образование  | URL: https://openedu.ru/ |
| 2. | Coursera | URL: https://www.coursera.org |

* 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 лабораторных и самостоятельных занятий по дисциплине оснащены ­­­­­­­­­­­­­­­­­­­­­­­­ ПЭВМ (операционная система, офисные программы), с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.