

МАШИННОЕ ОБУЧЕНИЕ В БИЗНЕСЕ

(прикладные Байесовские методы)

ФИО преподавателя: Кочуров Максим Вадимович

- Выпускник ЭФ МГУ 2018, СколТех 2020
- Разработчик PyMC
- Партнер PyMC Labs

Цель освоения дисциплины: практическое применение байесовских методов в исследованиях.

О курсе: данный курс нацелен освоение байесовских методов на практике. Продвинутое моделирование – это не сопряженные распределения (хотя и они порой оказываются полезными), а по большей части проецирование знаний о проблеме в код. Для успешного исследования также понадобятся навыки решения проблем сходимости, параметризации, выбора из нескольких моделей. Курс с фокусом на программирование на Python / PyMC. Используемые сервисы: Yandex Cloud, Yandex DataSphere.

Место дисциплины в учебном плане: дисциплина по выбору, 4 триместр

Входные требования для освоения дисциплины

Для успешного освоения данного курса требуются знания и навыки, полученные в следующих дисциплинах:

- Эконометрика (продвинутый уровень)
- Программирование, основы алгоритмов и анализ данных в Python
- Вероятностные модели (адаптационный курс)

Краткое содержание дисциплины:

- Байесовское мышление
- Принципы моделирования
- Программирование байесовских моделей в Python
- Валидирование модели
- НМС - решение жизненных проблем
- Интерпретируемые Гауссовские Процессы
- Кросс-валидация
- Сравнение байесовских моделей

Общая трудоемкость дисциплины составляет 3 зачетные единицы, 108 часов.

Язык преподавания: русский

Промежуточная аттестация: по дисциплине проводится в форме накопленной оценки:

60% домашние работы

40% командный проект (экзамен предусматривает защиту проекта)

Календарно-тематический план курса:

Тема 1. Байесовское мышление, вводная часть (4 ауд. часа)

- a. **Лекция** вводная
- b. **Семинар** введение в РуМС
 - i. Настройка окружения

Тема 2. Принципы моделирования (4 ауд. часа)

- c. **Лекция** по моделированию
 - i. Базовая модель
 - ii. Параметризация
 - iii. Выбор априорного распределения
 - iv. Проверка априорного распределения
 - v. НМС для практики
- d. **Семинар** по моделированию
 - i. Разбор кейса по составлению модели
 - ii. Иерархичность
 - iii. Матрица адаптации
 - iv. Проверка априорного распределения (практика)
 - v. «Житейские хитрости»
 - vi. Освоение сервиса Yandex DataSphere, выполнение практической работы в облаке

Тема 3. Parameter Recovery (4 ауд. часа)

- e. **Лекция** Байесовские ABCDE-тесты
- f. Продвинутый **Семинар** по Parameter Recovery
 - i. Когда надо делать Parameter Recovery
 - ii. Как правильно делать Parameter Recovery
 - iii. Введение в Dask
 - iv. Эффективный Хаггау
 - v. «Житейские хитрости»

Тема 4. Дебаг Байесовского вывода (4 ауд. часа)

- g. **Лекция** Вычислительные проблемы НМС
 - i. Обзор проблем на практике
 - ii. Обзор инструментов для решения проблем
- h. **Семинар** дебаг НМС
 - i. Разбор примера с проблемами
 - ii. Последовательность действий при дебаге НМС
 - iii. Выполнение практической работы в облаке

Тема 5. Временные ряды и Гауссовские процессы (8 ауд. часов)

- i. **Лекция** Введение в Гауссовские процессы
 - i. Разбор основных гиперпараметров
 - ii. Разбор примера
- j. **Семинар** Введение в Гауссовские процессы
 - i. Prior predictive для гауссовских процессов
 - ii. Выполнение практической работы в облаке
- k. **Лекция** Гауссовские процессы
 - i. Эффективная параметризация
 - ii. Учет сезонности и трендов
 - iii. Учет прочих факторов
- l. **Семинар** Кросс-валидация
 - i. Разбор кейса пор временной ряд
 - 1. Модель без тренда
 - 2. Модель с трендом
 - 3. Модель с трендом и сезонностью
 - ii. Сравнение моделей через LOO
 - iii. Построение прогноза
 - iv. Выполнение практической работы в облаке

БАЛЛЬНО-РЕЙТИНГОВАЯ СИСТЕМА

Оценка	Минимальное количество баллов	Максимальное количество баллов
<i>Отлично</i>	127,5	150
<i>Хорошо</i>	97,5	127,4
<i>Удовлетворительно</i>	60	97,4
<i>Неудовлетворительно</i>	—	59,9

Примечания:

В случае, если магистрант за триместр набирает менее 20% баллов от максимального количества по дисциплине, то уже на промежуточном контроле (и далее на пересдачах) действует следующее правило сдачи: «магистрант может получить только оценку «Удовлетворительно», и только если получит за промежуточный контроль, включающий весь материал дисциплины, не менее, чем 85% от баллов за промежуточный контроль».

Ресурсное обеспечение

Python / PyMC. Используемые сервисы: Yandex Cloud, Yandex DataSphere

Описание материально-технической базы

Для организации занятий по дисциплине необходимы следующие технические средства обучения: компьютерный класс с проектором.