

Введение в машинное обучение

Higher School of Economics

Автор: Константин Вячеславович Воронцов, Профессор, Евгений Соколов, Старший преподаватель

Платформа Coursera

Ссылка на курс: <https://ru.coursera.org/learn/vvedenie-mashinnoe-obuchenie>

Не так давно получил распространение термин «большие данные», обозначивший новую прикладную область — поиск способов автоматического быстрого анализа огромных объёмов разнородной информации. Наука о больших данных ещё только оформляется, но уже сейчас она очень востребована — и в будущем будет востребована только больше. С её помощью можно решать невероятные задачи: оценивать состояние печени по кардиограмме, предсказывать зарплату по описанию вакансии, предлагать пользователю музыку на основании его анкеты в интернете. Большими данными может оказаться что угодно: результаты научных экспериментов, логи банковских транзакций, метеорологические наблюдения, профили в социальных сетях — словом, всё, что может быть полезно проанализировать. Самым перспективным подходом к анализу больших данных считается применение машинного обучения — набора методов, благодаря которым компьютер может находить в массивах изначально неизвестные взаимосвязи и закономерности. На факультете компьютерных наук ВШЭ и в Школе анализа данных есть люди, активно использующие машинное обучение и разрабатывающие новые подходы к нему. Именно они — преподаватели этого курса. Вы изучите основные типы задач, решаемых с помощью машинного обучения — в основном речь пойдёт о классификации, регрессии и кластеризации. Узнаете об основных методах машинного обучения и их особенностях, научитесь оценивать качество моделей — и решать, подходит ли модель для решения конкретной задачи. Наконец, познакомитесь с современными библиотеками, в которых реализованы обсуждаемые модели и методы оценки их качества. Для работы мы будем использовать реальные данные из реальных задач.

Краткая программа курса:

Неделя 1. Введение. Примеры задач. Логические методы: решающие деревья и решающие леса.

Неделя 2. Метрические методы классификации. Линейные методы, стохастический градиент.

Неделя 3. Метод опорных векторов (SVM). Логистическая регрессия. Метрики качества классификации.

Неделя 4. Линейная регрессия. Понижение размерности, метод главных компонент.

Неделя 5. Композиции алгоритмов, градиентный бустинг. Нейронные сети.

Неделя 6. Кластеризация и визуализация. Частичное обучение.

Неделя 7. Прикладные задачи анализа данных: постановки и методы решения.

Слушателю нужно знать об основных понятиях математики: функциях, производных, векторах, матрицах. Для выполнения практических заданий потребуются базовые навыки программирования. Очень желательно знать Python. Задания рассчитаны на использование этого языка и его библиотек `numpy`, `pandas` и `scikit-learn`. Чтобы успешно завершить курс, нужно набрать проходную сумму баллов за тесты и практические задания, а также выполнить финальный проект, посвящённый решению прикладной задачи анализа данных. Этот курс будет полезен каждому, кто хочет постигнуть искусство предсказательного моделирования и освоить интеллектуальный анализ данных.