Курс: Data Science: Машинное обучение в R

Длительность: 40 ак. часов

О курсе

Данный курс предназначен для изучения алгоритмов машинного обучения с практическим применением техник машинного обучения реализованных в **R**. Рассматриваются понятия **data mining**, измерения производительности и уменьшения размерности, регрессионные модели, байессовская модель, **SVM** и ассоциативные правила для анализа. После успешного завершения данного курса вы сможете понимать и объяснять принципы работы алгоритмов машинного обучение, и применять данные алгоритмы на реальных задачах в **больших данных**.

Аудитория

Специалисты по работе с большими данными, бизнес аналитики и руководители желающие получить расширенную практическую и теоретическую подготовку по методам **Data Mining** для участия в проектах анализа больших данных и машинного обучения.

Предварительная подготовка

- Понимание основ статистики
- Опыт работы с **R-Studio** или знания в рамках курса **DSAV-Data Science: Аналитика и визуализация больших данных в R**

Программа

1. **Основы статистики и простая линейная регрессия**
   Что такое ваши данные. Статистические выводы. Введение в машинное обучение. Простая линейная регрессия. Диагностика и трансформация. Коэффициент определенности. Методы оценки моделей и производительности.

2. **Базовое программирование с R (опционально)**
   Введение в **R**. Что такое **R**? **R-Studio**, пакеты и рабочая область. Основные элементы языка **R**. Типы объектов данных. Введение функций и управляющих операторов. Функции. Программирование функций. Подключение библиотек в **R-Studio**.

mailto:sales@bigdataschool.ru
+7 (985) 162-29-63

3. **Подготовка данных (опционально)**

   Принципы формирование Dataset (набор данных). Локальный импорт / экспорт данных. Работа с отсутствующими данными (NA) . Категориальные данные. Формирование обучающего и тестового набора данных. Вопросы масштабирования и автоматизации. Препроцессинг данных.

4. **Линейная регрессия и обобщенная линейная модель**

   P-value - ошибки первого рода. Допущения и диагностика. Оценка максимального правдоподобия. Интерпретация модели. Оценка соответствия модели.

   Обобщенные линейные модели:

   - Простая линейная регрессия
   - Множественная линейная регрессия
   - Логистическая регрессия
   - Полиномиальная регрессия

   Метод опорных векторов (SVR) и деревья решений

   Деревья решений. **Bagging**. Случайные леса. **Boosting**. Важность переменной. Сортировка полей и поддержка векторного классификатора. Метод опорных векторов.

   Оценка производительности регрессионной модели. Коэффициенты линейной регрессии.

5. **Алгоритмы классификации.**

   Логистическая регрессия.

   Алгоритм ближайших соседей.

   Алгоритм K-ближайших соседей. Выбор K и меры расстояния.

   Наивный байессовский анализ и "проклятие размерности" Условная вероятность: теорема Байеса. Оценка Лапласа. Уменьшение размерности. Процедура PCA. Ridge и регрессия Лассо. Перекрестная проверка.

   Классификация с помощью деревьев решений.

   Классификация методом случайных деревьев.

   Оценка производительности классификационной модели.

6. **Кластерный анализ**

   Кластерный анализ.

   K-means кластеризация

   - Выбор количества кластеров
   - Типовые ошибки при кластеризации

   Иерархическая кластеризация. Принципы построения дендрограмм.

7. **Ассоциативные правила**

   Правила Априори алгоритма

   Основные приниципы и построение модели в R

8. **Машинные алгоритмы с переобучением (Reinforcement learning)**

   Верхняя граница достоверности (UCB - Upper Confidence Bound)

Примеры по Томпсону
Сравнение алгоритмов
Реализация алгоритмов в К

9. NLP алгоритмы (Алгоритмы текстовой обработки)

Основы Natural Language Proccesing

10. Глубокое Обучение (Deep Learning)

Отличие машинного обучение(Machine Learning) от глубокого обучения (Deep Learning)
Искуственные Нейронные Сети (Artificial Neural Networks) :

- План атаки
- Нейроны
- Активация нейронов
- Как работают нейронные сети и персептроны
- Сигмоидные нейроны
- Сетевая топология и скрытые функции
- Метод обратного распространения ошибки с градиентным спуском