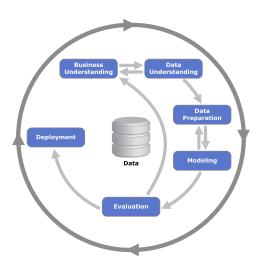


Лекция 11 Data Mining в реальных системах

Николай Анохин

22 мая 2015 г.



План занятия

Работа с признаками

Алгоритмы машинного обучения

Особенности реальных систем

Что дальше

Работа с признаками

Конструирование признаков

Преобразование признаков

- Дискретизация
- Проекции
 - ► PCA
 - Random projections
- Заполнение отсутствующих значений
- Удаление шума

Отбор признаков

Как "нерелевантные", так и "релевантные" признаки могут быть вредными

- 1. Независимо от алгоритма обучения: backward elimnation, forward selection
 - mutual information
 - коэффициент корреляции
 - линейная модель
 - генетические алгоритмы
- 2. С использованием алгоритма обучения (кросс-валидация, hold-out)

Алгоритмы машинного обучения

Особенности реальных систем

Что дальше

Вопросы

