

Eksploracja danych i uczenie maszynowe

Przygotowanie danych

Uczenie maszynowe jako element eksperymentów *data science*

Data science jako metoda naukowa

Eksploracja danych jako technika wspomagania decyzji

Modelowanie

Wiedza i proces uczenia

Przykładowy eksperyment *data science*

Ocena przydatności danych

Dane źródłowe

Zmienne

Reprezentatywność danych

Szeregi czasowe

Wstępne przetwarzanie danych

Uzupełnianie brakujących danych

Poprawianie błędnych danych w zmiennych numerycznych

Poprawianie błędnych danych w zmiennych kategoriowych

Wygładzanie szeregów czasowych

Przetwarzanie wyrażeń języka naturalnego

Redukcja wymiarów

Wzbogacanie danych

Równoważenie danych

Zmienne wyliczeniowe

Zastąpienie zmiennych wspólnym rozkładem prawdopodobieństwa

Wydzielenie danych testowych

Wzorzec eksperymentu Data Science

Modele predykcyjne

Klasyfikacja

Klasyfikacja poprzez indukcję drzew decyzyjnych

Pojedyncze drzewa decyzyjne

Kombinacje drzew decyzyjnych

Klasyfikacja z użyciem maszyny wektorów nośnych

Przetwarzanie języka naturalnego

Klasyfikacja probabilistyczna

Klasyfikatory binarne a klasyfikacja wieloklasowa

Regresja

Model regresji wielorakiej
Zmienne kateryczne w modelach regresji
Regresja poprzez indukcję drzew decyzyjnych
Sztuczne sieci neuronowe
Głębokie sieci neuronowe
Grupowanie
Analiza skupień
Algorytmy grupowania
Grupowanie w celu znajdowania podobnych obiektów
Grupowanie w celu kompresji
Wykrywanie anomalii
Rekomendowanie
Systemy rekomendujące
Odkrywanie asocjacji
Rekomendowanie przez filtrowanie kolektywne
Rekomendowanie hybrydowe
Prognozowanie
Szeregi czasowe
Naiwne metody prognozowania
Modele średniej ważonej
Modele ARIMA
Modele nieliniowe

Ocena i poprawa jakości modeli predykcyjnych

Reguła powrotu do średniej
Kryteria oceny modeli eksploracji danych
Ocena dokładności predykcji i walidacja krzyżowa
Ocena jakości modeli klasyfikacyjnych
Macierz pomyłek
Krzywa ROC
Wykresy precyzja/czułość i zysku
Trafność klasyfikacji
Klasyfikatory wieloklasowe
Ocena jakości modeli regresyjnych
Miary oceny modeli
Ocena jakości modeli grupujących
Ocena jakości modeli rekomendujących
Ocena jakości modeli prognozujących

Porównanie jakości modeli

Poprawa jakości modeli

Automatyczna poprawa jakości modeli uczenia nadzorowanego

Znalezienie optymalnej liczby klastrów

Publikacja modeli predykcyjnych jako usług WWW

Wzorcowy eksperyment *data science*

Predykcyjne usługi WWW

Zapytania predykcyjne ad-hoc

Wsadowe zapytania predykcyjne