

Machine Learning – case study

Wykonał: Maciej Niemiec

Numer albumu: 107162

Spis treści

1. Uczenie Maszynowe ze Wzmocnieniem (Reinforcement Learning - RL)	3
2. Uczenie Maszynowe z Nauczycielem (Supervised Learning)	3
3. Uczenie Maszynowe Nienadzorowane (Unsupervised Learning)	4

1. Uczenie Maszynowe ze Wzmocnieniem (Reinforcement Learning - RL)

Uczenie maszynowe ze wzmocnieniem (ang. *reinforcement learning*, *RL*) to rodzaj uczenia, w którym **agent (model) uczy się podejmowania decyzji poprzez interakcję z otoczeniem w celu maksymalizacji długoterminowej nagrody**. Różni się ono od innych rodzajów uczenia, ponieważ **nie dostarcza gotowych odpowiedzi (etykiet)**, lecz skupia się na eksploracji i optymalizacji strategii działania.

Kluczowe elementy RL:

- **Agent:** System uczący się, np. robot lub program grający w szachy.
- **Otoczenie:** Świat, w którym agent działa, np. plansza gry czy symulacja.
- **Stan (State):** Opis sytuacji w otoczeniu w danym momencie.
- **Akcja (Action):** Działanie, które agent może podjąć w danym stanie.
- **Nagroda (Reward):** Informacja zwrotna, która ocenia skuteczność akcji agenta.
- **Polityka (Policy):** Strategia, która mówi agentowi, jakie akcje wybierać w danych stanach.
- **Funkcja wartości (Value Function):** Ocenia, jak korzystny jest dany stan lub akcja w kontekście przyszłych nagród.

Przykład zastosowania: Agent może uczyć się grać w grę wyścigową. Za każdym razem, gdy wypadnie z zakrętu, otrzymuje "karę" (mniejszą wagę na dane zachowanie, które doprowadziło do błędu). W przypadku, gdy dojedzie do mety i osiągnie lepszy czas niż poprzednio, do jego zachowań z danego podejścia przypisuje się dodatkowe wagi.

2. Uczenie Maszynowe z Nauczycielem (Supervised Learning)

Uczenie maszynowe z nauczycielem (ang. *supervised learning*) polega na **trenowaniu modelu na zbiorze danych, który zawiera zarówno dane wejściowe (cechy), jak i znane odpowiedzi (etykiety)**. Model uczy się na podstawie tych danych, aby przewidywać etykiety dla nowych, niewidzianych wcześniej danych.

Kluczowe aspekty:

- **Etykieta:** Informacja, którą model próbuje przewidzieć, np. "jabłko".
- **Cechy:** Dane wejściowe, np. kolor, kształt.
- **Klasyfikacja:** Jednym z zadań w uczeniu z nauczycielem jest klasyfikacja, czyli przypisywanie danych do znanych klas.

Przykład zastosowania: Aplikacja do rozpoznawania owoców. Model jest trenowany na zdjęciach owoców (np. jabłek i bananów) z przypisanymi do nich etykietami ("jabłko", "banan"). Model analizuje te dane i uczy się rozróżniać owoce na podstawie cech takich jak kształt, kolor czy rozmiar. Innym przykładem zastosowania jest przewidywanie cen domów na podstawie cech takich jak powierzchnia, liczba pokoi i lokalizacja, co jest możliwe dzięki Perceptronowi Wielowarstwowemu (MLP), który znajduje zastosowanie w klasyfikacji i regresji.

3. Uczenie Maszynowe Nienadzorowane (Unsupervised Learning)

Uczenie nienadzorowane to typ uczenia maszynowego, w którym algorytm **sam odkrywa naturalne skupiska w danych**, bez wcześniejszej wiedzy o etykietach. Głównym celem jest **grupowanie danych w taki sposób, aby punkty w tej samej grupie (klastrze) były do siebie bardziej podobne niż do punktów w innych grupach**.

Przykładowe algorytmy klasteryzacji (grupowania):

- **K-means:**
 - Opis: Dzieli dane na k klastrow na podstawie podobieństwa. Jest **szybki i prosty w implementacji**.
 - Działanie: Użytkownik wybiera liczbę klastrow k , a algorytm losowo umieszcza centra klastrow. Następnie iteracyjnie przypisuje każdą daną do najbliższego centrum i oblicza średnią odległość od centrum dla każdego punktu w klastrze, powtarzając proces aż do satysfakcjonującej średniej odległości.
 - Wady: **Musi mieć wybraną liczbę klastrow k z góry**. Ma tendencję do znajdowania klastrow o **określonym, kulistym kształcie** i **nie jest wrażliwy na szumy** (zawsze przypisuje punkt do klastra).
 - Zastosowanie: Często wykorzystywany do **segmentacji klientów sklepu na podstawie zachowań zakupowych**.
- **Hierarchiczna klasteryzacja:**
 - Opis: Tworzy hierarchię klastrow, którą można wizualizować w formie dendrogramu.
 - Zalety: **Nie wymaga podania liczby klastrow z góry**.
 - Wady: **Duża złożoność obliczeniowa** przy dużych zbiorach danych.
 - Zastosowanie: Analiza taksonomii biologicznej.
- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise):**
 - Opis: Algorytm, który **świetnie znajduje skupiska punktów** i identyfikuje punkty szumu.
 - Działanie: Opiera się na dwóch parametrach: ϵ (epsilon - jak blisko muszą być punkty) i MinPts (minimalna liczba punktów tworząca skupisko). Dzieli punkty na: rdzeniowe, brzegowe i szum.
 - Zalety: **Nie wymaga z góry określonej liczby klastrow** i ignoruje pojedyncze, odizolowane lokale.
 - Zastosowanie: Analiza lokalizacji kawiarni w mieście, automatyczne znajdowanie popularnych ulic z kawiarniami.