

## Popularnonaukowe streszczenie projektu

Projekt poświęcony jest statystycznemu uczeniu się w oparciu o dane wysokiego wymiaru. Dane wysokiego wymiaru dotyczą przypadków, gdy rozmiar próbki  $n$  jest duży, liczba parametrów modelu  $p$  jest większa niż  $n$  lub zarówno  $n$ , jak i  $p$  są duże. Obecnie takie zbiory danych można często spotkać w biologii lub genetyce. Jest to problem bardzo ambitny i zwykle klasyczne narzędzia statystyczne nie są odpowiednie do rozwiązania tych problemów. W sytuacji, gdy rozmiar próbki jest duży klasyczne narzędzia statystyczne zwykle prowadzą do dużej złożoności obliczeniowej. Gdy liczba parametrów jest większa niż rozmiar próbki klasyczne narzędzia statystyczne prowadzą do niewłaściwie postawionych zagadnień. W ostatnich latach w społecznościach zajmujących się statystyką oraz uczeniem maszynowym wiele uwagi poświęca się statystycznemu uczeniu się w oparciu o dane wysokiego wymiaru. Pomimo znaczących postępów na tym polu, nadal znajdziemy obszary wymagające zbadania i niniejszy projekt podejmuje niektóre z nich. Dokładniej, skupimy się na następujących zagadnieniach:

- Analiza metod aproksymacji stochastycznej w problemach niegładkich i niewypukłych,
- Algorytmy Langevina w oparciu o metody Monte Carlo do wnioskowania bayesowskiego dla dużych danych,
- Badanie struktury modeli grafowych.

W projekcie planujemy skonstruować algorytmy będące w stanie rozwiązywać problemy związane z powyższymi zagadnieniami. Ponadto będziemy dokładnie badać własności tych procedur, na przykład ich tempo zbieżności czy własności statystyczne.

Pole możliwych zastosowań otrzymanych wyników jest bardzo szerokie, na przykład: kategoryzacja tekstu, multimedia (rozpoznawanie twarzy, obiektów), biologia (predykcja genów, predykcja struktury białek), chemia (odkrycia farmakologiczne) czy medycyna.