

Szczecin, 19.09.2024 r.

Streszczenie rozprawy doktorskiej

Uczenie modeli liniowych z regularyzacją dla małych zbiorów danych

Autor: mgr inż. Jacek Klimaszewski

Promotor: dr hab. inż. Marcin Korzeń, prof. ZUT

W niniejszej rozprawie poruszono zagadnienie uczenia modeli liniowych z regularyzacją, skupiając się na małych zbiorach danych (liczba atrybutów znacznie przewyższa liczbę obserwacji). Przeanalizowano i zbadano wydajność istniejących procedur uczenia na tych zbiorach oraz różnych form regularyzacji modeli. Sprawdzone aspekty numeryczne algorytmów, m.in. wpływ typu zmiennoprzecinkowego pojedynczej i podwójnej precyzji oraz rodzaju minimalizacji kierunkowej na szybkość uczenia oraz sposoby reprezentacji danych uczących. Wykazano teoretycznie i empirycznie, że w pewnym zakresie parametrów wykorzystanie formuły Woodbury'ego bądź rozkładu QR znacznie skraca czas uczenia modeli z regularyzacją ℓ^2 — sprawdzono analogiczne rozwiązanie dla modeli wyposażonych w rzadkie funkcje kary. Eksperymenty wykonano w środowisku Python wraz z fragmentami zaimplementowanymi w języku C++, biorąc za punkt odniesienia istniejące biblioteki uczenia maszynowego, jak scikit-learn czy liblinear. Przetestowano działanie proponowanych rozwiązań w zadaniach klasyfikacji na rzeczywistych i syntetycznych zbiorach danych oraz wskazano potencjalne zastosowania uczenia regularyzowanych modeli liniowych w problemach nieliniowych.

słowa kluczowe: uczenie maszynowe, modele liniowe, regularyzacja, optymalizacja, metoda Newtona

19.09.2024

Jacek Klimaszewski