Założenia dla projektu realizowanego w ramach przedmiotu: ‘Metody optymalizacji w analizie danych’.

W ramach projektu zespoły (maksymalnie dwuosobowe) będą miały do rozwiązania zadanie optymalizacji z następującego zakresu analizy danych:

1. Budowa modelu regresji liniowej z wykorzystaniem modelu LASSO interpretowanego jako zadanie mieszane optymalizacji całkowitoliczbowej
2. Budowa modelu regresji liniowej z wykorzystaniem modelu LASSO interpretowanego jako zadanie optymalizacji nieróżniczkowalnej
3. Budowa modelu predykcyjnego dla klasyfikacji danych z wykorzystaniem modelu SVM interpretowanego jako zadanie optymalizacji nieróżniczkowalnej
4. Budowa modelu predykcyjnego dla klasyfikacji danych z wykorzystaniem modelu SVM interpretowanego jako zadanie kwadratowe optymalizacji
5. Budowa modelu predykcyjnego dla klasyfikacji danych z wykorzystaniem modelu SVM interpretowanego jako zadanie optymalizacji całkowitoliczbowej
6. Budowa modelu nieliniowego regresji traktowanego jako nieliniowe zadanie optymalizacji

W ramach projektu wykorzystywane będą następujące algorytmy rozwiązywania zadań 1)-6):

1. Metody podziału i ograniczeń z funkcją ‘warm start’ wykorzystującą przybliżenie rozwiązania za pomocą metody pierwszego rzędu
2. Metody płaszczyzn odcinających
3. Metody pierwszego rzędu typu przyrostowego (‘incremental’)
4. Metody subgradientowej dla wypukłych zadań optymalizacji
5. Metody pierwszego rzędu wykorzystujące składnik proksymalny (‘proxy term’)
6. Metody stochastycznej pierwszego rzędu dla wypukłych zadań optymalizacji
7. Metody zmiennej metryki o ograniczonej pamięci dla zadań z ograniczeniami kostkowymi
8. Metody Newtona typu obszaru zaufania dla zadań z ograniczeniami kostkowymi.

Projekty będą realizowane z wykorzystaniem przede wszystkim środowiska języka R w oparciu o pakiety: ROI – R Optimization Infrastructure; CVXR – pakiet oparty na obiektowo-zorientowanym języku dla optymalizacji wypukłej stosującej reguły DCP (Disciplined Convex Programming).

Projekty wykorzystujące metody optymalizacji całkowitoliczbowej mogą również wykorzystywać środowisko obliczeniowe optymalizacji IBM – ILOG CPLEX Optimization Studio.

Projekty będą odwoływały się do ogólnie dostępnych danych takich jak ‘Diabetes’, czy ‘Leukemia’ jak również do danych generowanych na potrzeby weryfikacji skuteczności testowanych metod.