

Wnioskowanie statystyczne w przypadku danych o wielkim wymiarze: wyzwanie niesione przez nauki o życiu i nie tylko

Jacek Koronacki

Institut Podstaw Informatyki PAN, Warszawa

Miniony wiek bywa nazywany wiekiem informacji. Moce obliczeniowe komputerów oraz pojemności ich pamięci rosły w ostatnich dziesięcioleciach nieomal z dnia na dzień. Dziś nierzadko słyszymy, iż weszliśmy w wiek biologii a nawet szerzej, wiek nauk o życiu. Stało się tak między innymi dlatego, że niezwykle rozwinęły się biotechnologie pozwalające na zbieranie masowych danych o żywych komórkach. Wyzwanie, jakie się za tym kryje, bierze się stąd, że typowe dane w takich zastosowaniach biologicznych charakteryzują się małą liczbą obserwacji (obiektów) – np. rzędu dziesiątków lub setek – z których każda opisana jest tysiącami lub większą liczbą atrybutów (cech). Dobrymi i ważnymi przykładami problemów tego typu są dane mikromacierzowe, dotyczące tzw. ekspresji genów, lub dane proteomiczne. W ramach wykładu przedstawiony zostanie stan badań w obszarze uczenia pod nadzorem w opisanej sytuacji.