

Johnson-Lindenstrauss típusú garanciák és általánosítási korlátok eltérő komplexitású halmazokra

Ata Kabán, Henry Reeve és Jakramate Bootkrajang

University of Birmingham
School of Computer Science

A.Kaban@bham.ac.uk

Azt vizsgáljuk, hogyan vezethetők le nagy valószínűséggel, egyidejűleg érvényes Johnson–Lindenstrauss-típusú normamegőrzési és statisztikai tanulásbeli általánosítási korlátok, amikor a vetítendő geometriai halmaz, illetve a tanulási feladat hipotézisosztálya különböző komplexitású komponensek konvex kombinációit tartalmazza. A hagyományos korlátok jellemzően a legnagyobb komplexitású összetevőtől függenek, ezért az alacsonyabb komplexitású részekre szükségtelenül laza korlátokat adhatnak.

Legyen $S_1, \dots, S_m \subset \mathbb{R}^d$ a d -dimenziós euklideszi tér korlátos részhalmazainak családja, és legyen $S = \text{conv} \left(\bigcup_{j=1}^m S_j \right)$. Bevezetjük az S elemeinek elemenkénti komplexitását, amelyet az adott elem összes lehetséges konvex előállításában szereplő halmazok Gauss-szélességeiből képzett súlyozott átlagok infimumaként értelmezünk. Erre a fogalomra építve felső korlátokat vezetünk le az elemenkénti komplexitás alapján korlátozott részhalmazok Gauss- és Rademacher-szélességére. A módszert két területen alkalmazzuk:

(1) A véletlen vetítésen alapuló dimenziócsökkentésben az euklideszi normák és távolságok egyidejű megőrzésére vezetünk le új, adaptív korlátot, ami az egyes pontok elemenkénti komplexitásától függ, miközben a komponenshalmazok maximális Gauss-szélessége csak logaritmikus tagban jelenik meg. Ez szorosabb normamegőrzési garanciákat tesz lehetővé az alacsony komplexitású régiókban.

(2) A statisztikai tanulásban eltérő komplexitású függvényosztályok súlyozott együtteseire vezetünk le általánosítási korlátokat, amelyek az egyes előrejelzők elemenkénti Rademacher-komplexitásától függenek. Eredményeink elméleti megalapozást adnak heterogén együttes tanulási módszerekhez, és útmutatást nyújtanak azok kialakításához.

Az elméleti eredményeket regularizált, boosting típusú, heterogén együttes tanulási algoritmusokkal végzett kísérletekkel szemléltetjük. A módszerek a vizsgált osztályozási feladatokban versenyképes empirikus teljesítményt mutatnak.

Az előadás az [1]-ben hivatkozott tanulmányon alapul.

Hivatkozások

- [1] Reeve, H., Kabán, A. & Bootkrajang, J. Heterogeneous sets in dimensionality reduction and ensemble learning. Machine Learning 113, 1683–1704 (2024). <https://doi.org/10.1007/s10994-022-06254-0>