

單位元高斯過程最佳分割的次層搜尋法

黃思緯

清華大學統計與數據科學研究所

摘要

高斯過程為一種靈活且具統計基礎的半參數局部平滑化方法。但在資料量有限時，其表現高度依賴核函數的設計。對比於文獻中常見的相乘型核函數，加法型核函數雖能捕捉變數間各種交互作用，但兩者分別有彈性不足與易過度配適的問題。為此，Unital Gaussian Process (UGP) 透過對變數集合進行分割，建構出混有相乘及加法結構特性的彈性核函數，以提升模型對複雜資料之適應能力。然而，UGP 的效能高度依賴所挑選的分割，而尋找最適分割的過程，會面臨分割數量隨變數個數增加而呈階乘級上升，與不同分割的 UGP 模型間缺乏包含關係等挑戰。本文基於分割間「較細」與「較粗」關係所建立之階層式結構，提出一套最適分割搜尋法。其以最細分割為起點，利用對次層的分割進行分類，從而有效搜尋最適分割。透過模擬研究，我們發現此方法不僅能大量節省窮舉搜尋所帶來的高昂計算成本，準確搜尋出最適分割，同時亦可維持優異的預測效能。

關鍵字：相加型核、階層式結構、核函數、模型選擇、類神經網路、相乘型核