

子集抽樣：大數據分析的省力神器

張明中 副研究員

中央研究院統計科學研究所

隨著數據量的增加，其所蘊含的資訊也越來越豐富。然而，這些資訊並不總是能夠被數據分析的技術輕易擷取。面對大規模數據挑戰時，電腦可能會因負荷過重而無法運行，或者即使能夠運行，也可能會遇到演算法無法收斂，從而無法獲得最優解的問題。在本次講座中，我們將探討子集抽樣 (Subsampling) 這一技術，它不僅易於理解且容易實施，更專門針對從龐大數據集中抽取代表性的資訊。子集抽樣技術顯示出解決這些問題的巨大潛力，並能幫助我們揭開被大數據所遮蓋的寶藏。