
SHOFIQL ISLAM, McMaster University

Comparer la performance des composantes principales linéaires et non-linéaires dans le contexte de l'intégration des données génomiques de grande dimension

Nous intégrons phénotype, expression des gènes et données d'expression miARN pour classer la tumeur ou la mort de patients atteint du cancer du poumon. Nous appliquons une analyse en composantes principales (ACP) linéaire et de noyau afin de réduire la dimension et d'intégrer ensuite trois domaines en se basant sur plusieurs approches non linéaires, y compris la régression logistique. Nous avons également effectué une vaste simulation pour comparer les performances. Nous observons qu'un nombre réduit de composantes principales de noyau renferme plus d'information sur un processus génétique, mais a échoué, dans ce cas, à montrer une meilleure capacité prédictive. À cette fin, il semble suffisant de réduire la dimension avec une ACP linéaire et un modèle de régression logistique.