Date: 15 June (Tue)
Time: 3-4 pm
Venue: G92 Madsen Bldg
Speaker: Rob Munro
Title: An Integrated Approach to Mining Data Streams

Abstract:
Processor speed may be doubling every 18 months, but the total amount of data we store is doubling in half that time, while IO speeds remain near constant. Therefore, learning from data streams has quickly become one of the most active fields in data mining and machine learning.

Mining highly correlated data has always been known as a difficult task in data mining, with ‘Naïve’ Bayes getting its name for its inability to learn correlations (known as the attribute independence assumption). Current state-of-the-art learners weaken this to capture pairwise correlations with quadratic efficiency.

This talk will present a recent advance in mining correlations that integrates techniques from the fields of outlier detection, clustering and supervised learning. The algorithm presented is able to learn and exploit correlations between a potentially unbounded number of attributes in one pass over an unbounded amount of data, with strictly bounded memory and processing costs.