Bregman Divergences for Exploratory Data Analysis

Colin Fyfe

The University of the West of Scotland

We discuss the family of Bregman divergences and show links between the members of this family and the exponential families of distributions. Specifically we show that certain divergences are optimal for clustering samples from these distributions. We show how these divergences may be used to create topology preserving mappings which optimally capture the manifold on which the data lies. We also apply the divergences to create a Pseudo Metric Multidimensional Scaling which has the happy proprerty (like the Sammon Mapping) of discounting data which are far apart. Finally we show how Bregman divergences can be used to create projections of single and dual stream data.