Clustering of Sun Exposure Measurements

In a medically motivated Sun-exposure study, questionnaires concerning Sun-habits were collected from a number of subjects together with UV radiation measurements. This paper focuses on identifying clusters in the heterogeneous set of data for the purpose of understanding possible relations between Sun-habits exposure and eventually assessing the risk of skin cancer. A general probabilistic framework originally developed for text and Web mining is demonstrated to be useful for clustering of behavioral data. The framework combines principal component subspace projection with probabilistic clustering based on the generalizable Gaussian mixture model.