Graph Regularized Feature Selection with Data Reconstruction

Feature selection is a challenging problem for high dimensional data processing, which arises in many real applications such as data mining, information retrieval, and pattern recognition. In this paper, we study the problem of unsupervised feature selection. The problem is challenging due to the lack of label information to guide feature selection. We formulate the problem of unsupervised feature selection from the viewpoint of graph regularized data reconstruction. The underlying idea is that the selected features not only preserve the local structure of the original data space via graph regularization, but also approximately reconstruct each data point via linear combination. Therefore, the graph regularized data reconstruction error becomes a natural criterion for measuring the quality of the selected features. By minimizing the reconstruction error, we are able to select the features that best preserve both the similarity and discriminant information in the original data. We then develop an efficient gradient algorithm to solve the corresponding optimization problem. We evaluate the performance of our proposed algorithm on text clustering. The extensive experiments demonstrate the effectiveness of our proposed approach.