Paper Review: Deep Subspace Clustering Networks

Brief Summary:

Motivated by the presence of non-linear subspaces in various computer vision related tasks that make use of subspace clustering such as face image clustering, the authors propose an auto-encoder based architecture for unsupervised non-linear subspace clustering. The 2-step process for subspace clustering involves estimating an affinity matrix to estimate the affinity for every pair of data points followed by spectral clustering performed using the affinity matrix. To this end, the authors introduce a fully-connected layer without bias and non-linear activations at the encoder-decoder junction to estimate the affinity matrix to which spectral clustering is applied. The autoencoder that maps the data into a latent space introduces the non-linearity component.

In introducing the fully-connected layer at the encoder-decoder junction, the authors make use of the self-expressiveness property which states that for data drawn from multiple linear subspaces, each point in a particular subspace can be expressed as a linear combination of all other points in the same subspace. This property can be expressed as ($\boldsymbol{X} = \boldsymbol{XC}$) where $\boldsymbol{X}$ represents the data point matrix for a subspace and $\boldsymbol{C}$ represents the self-representation coefficient matrix. Building on previous work, the authors frame structural constraints on $\boldsymbol{C}$ as an optimization problem expressed as $min_C \|C\|_p + \frac{\lambda}{2}\|X - XC\|_F^2 \; s.t. (diag(C) = 0)$. In using the full-connected layer, the authors formulate $\boldsymbol{C}$ to be the weights of the FC-layer and $\boldsymbol{X}$ to be the output of the encoder layer (i.e: the latent space representation of the data). The optimization problem introduced before is then included in the loss-function to be simply solved during backpropagation.

The authors initially pretrain the autoencoder model on the limited available data without introducing the FC layer and then pass the entire data as a batch along with the FC layer to fine tune its weight parameters using gradient descent. The authors evaluate their model on 4 standard CV related subspace clustering datasets and compare the performance of their model against SOTA baselines to prove superior performance of their approach.

Paper Contributions:

Two of the most important contributions of the paper are the use of autoencoders to introduce non-linearity and the introduction of the FC layer to make use of the “self-expressiveness” property in order to explicitly estimate the affinity matrix. The secondary contributions of pretraining and fine-tuning steps introduced by the authors also help improve model performance over the previous SOTA methods.

Additional Comments:

The authors were able to show successful experimentation for subspace clustering on computer vision related problems, however, it might be of interest to apply a similar methodology in obtaining subspace clusters in other domains and to compare the performance with other supervised clustering methods. Experimentation could also be conducted to compare the computational complexity when utilising the l_1 vs the l_2 norm minimization on $\boldsymbol{C}$.