Homework #3 Solutions

7.2 $\mathcal{E}$ is an image with all zeros and with a single one at the center. We convolve this with the kernel

$$H_{ij} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{[(i-k-1)^2 + (j-k-1)^2]}{2\sigma^2}\right).$$

Convolution with the kernel is a dot product

$$\begin{aligned} R_{ij} &= \sum_{u,v} H_{i-u,j-v}\mathcal{E}_{uv} \\ &= \sum_{u,v} H_{i-u,j-v}\delta_{uv} \\ &= H_{ij} \end{aligned}$$

which is a circularly symmetric fuzzy blob.

14.1

a) The scalar quantity $(\mathbf{x}_j - \mathbf{c}_i)^T S(\mathbf{x}_j - \mathbf{c}_i)$ redefines the distance between the feature vector $\mathbf{x}_j$ to $\mathbf{c}_i$ from the Euclidean distance $\|\mathbf{x}_j - \mathbf{c}_i\|^2 = (\mathbf{x}_j - \mathbf{c}_i)^T(\mathbf{x}_j - \mathbf{c}_i)$ to $(\mathbf{x}_j - \mathbf{c}_i)^T S(\mathbf{x}_j - \mathbf{c}_i)$. Consequently, the nearest neighbor measure in k-means needs to use the new distance measure rather than the older Euclidean distance. Everything else remains the same.

b) One since only the distance measure has changed.

c) Each step in k-means reduces the value of the objective function. To see this, define the objective function as

$$E(M, \mathbf{c}) = \sum_{ij} M_{ij}(\mathbf{x}_j - \mathbf{c}_i)^T S(\mathbf{x}_j - \mathbf{c}_i).$$

In k-means, we set M_{ij} to

$$M_{ij} = 1, \text{ for } i = \min_{i'} (\mathbf{x}_j - \mathbf{c}_{i'})^T S(\mathbf{x}_j - \mathbf{c}_{i'}).$$

This is a local optimum for M. The update for $\mathbf{c}$ is

$$\mathbf{c}_i = \frac{\sum_j M_{ij}\mathbf{x}_j}{\sum_j M_{ij}}$$

which is the optimum solution for $\mathbf{c}$ while keeping M fixed. Each step lowers the value of $E(M, \mathbf{c})$ and therefore this should converge to a local optimum. [However, this is not a proof. In general, merely showing that each step reduces the objective does not guarantee convergence to a local optimum. The objective could reach a stable value while M and $\mathbf{c}$ oscillate.]

14.2 **Shi and Malik (2000):** In class, we worked out that

$$y^T(D - W)y \propto \text{cut}(A, B).$$

Similarly, $y^T Dy$ can be expressed as

$$\begin{aligned} y^T Dy &= \sum_i D_{ii} y_i^2 = \sum_{i \in A} D_{ii} + b^2 \sum_{i \in B} D_{ii} \\ &= \sum_{i \in A} (\sum_{j \in A} W_{ij} + \sum_{j \in B} W_{ij}) + \sum_{j \in B} (\sum_{i \in A} W_{ij} + \sum_{j \in B} W_{ij}) \\ &= \text{assoc}(A, V) + \text{assoc}(B, V) \end{aligned}$$

Consequently

$$\frac{2y^T(D-A)y}{y^T Dy} = \frac{2\text{cut}(A,B)}{\text{assoc}(A,V) + \text{assoc}(B,V)} \leq \frac{\text{cut}(A,B)(\text{assoc}(A,V) + \text{assoc}(B,V))}{\text{assoc}(A,V)assoc(B,V)}$$

since $\frac{x+y}{2} \geq \frac{xy}{x+y}$. From this perspective, we end up minimizing a lower bound of the original desired cost function. A more detailed proof is available in the Shi and Malik paper.

14.3 a) The strategies are obviously not equivalent. In one case, we recursively partition the clusters. In the other, we do not change the problem. Instead, we look at other eigenvectors. If we removed the cluster and then derived a new affinity measure—that would be much more similar to the recursive approach.

b) The eigenvector corresponding to the largest eigenvalue represents the dominant cluster. Thresholding the eigenvector gives a single cluster.

c) The second eigenvector represents the second dominant cluster.

d) The two strategies are similar if we first choose elements belonging to the dominant cluster from the first eigenvector and then continue with the sevond eigenvector and so on. The procedures give different results if we decide on cluster membership using all eigenvectors in parallel. The parallel approach gives different results from the recursive approach.