

1. (20) We have discussed two clustering algorithms, kmeans and winner-take-all.
 - a) (10 pts) What is the difference in terms of learning procedure between these two algorithms?
 - b) (10 pts) Although they differ in the learning procedure, they have the common goal (or objective function). Explain what they are trying to maximize/minimize?
2. (50 pts) A 2D dataset has four samples: A(0, 1.1), B(0, 0), C(1, 0), and D(1, 0.9). Using agglomerative hierarchical clustering and Manhattan city block distance as distance metric between two samples.
 - a) (20 pts) Show detailed steps of clustering using d_{max} as distance metric between two clusters
 - b) (20 pts) Show detailed steps of clustering using d_{min} as distance metric
 - c) (10 pts) Comment on the impact of using different distance metrics. What kind of cluster shape tends to be generated using these two distance metrics?
3. (30 pts) Read Hinton's t-SNE paper and the demo (links provided in the lecture note) and visualize the pima.tr dataset using a 2-D plot. Try both PCA and t-SNE. Both PCA and t-SNE are unsupervised dimensionality reduction approaches, but PCA is linear and t-SNE is nonlinear. Comment on other differences from these two methods from the perspectives of objective function, procedure, efficiency, and when to use what. Note: You need to implement PCA yourself but you can call t-SNE routine.