**M532** **Spring 2018**
**Problem Set 4.** Due Tuesday, April 3

# Part A: Theory and Foundations

## Problem 1

Prove that the competitive learning update rule moves the winning center closer to the data point.

## Problem 2

Show that competitive learning forgets when the learning rate is contant, i.e., $\epsilon = \epsilon_0$.

## Problem 3

This problem concerns the training of a specific center in $k$-means. Use induction to prove that

$$c^k = \frac{1}{k}\sum_{i=1}^{k} x^{(\mu_i)}$$

Here we assume that the sequence of points $\{x^{(\mu_1)}, x^{(\mu_2)}, \ldots, x^{(\mu_k)}\}$ all have $c$ as the nearest center. We use the supoersciprt $m$ to denote the $m$th time that $c$ is updated.

## Problem 4

In this problem we fill in the details of the proof that MDS provides the best approximate configuration to a non-Euclidean distance matrix. Let $D \in \mathbb{R}^{n\times n}$ be a non-euclidean distance matrix and $B = HAH$ with eigenvalues $\lambda_1 \geq \lambda_k > 0, \lambda_{k+1} = 0$ and $0 > \lambda_{k+2} \leq \ldots \leq \lambda_n$. (As usual $A_{ij} = -D_{ij}^2/2$ and $H$ is the usual centering matrix.) Let $\hat{B}$ be the best approximation to $B$ in the sense that

$$\hat{B} = \arg\min_C \sum_{ij}(B_{ij} - C_{ij})^2$$

where $C$ is taken from the set of symmetric positive semi-definite matrices.

a) Show that

$$\sum_{ij}(B_{ij} - C_{ij})^2 = trace((B - C)^2)$$

b) Using the spectral theorem we can write $S^T BS = \Lambda$ and $R^T CR = \hat{\Lambda}$. What can you say about the properties of the matrices $R, S$ and the signs of the entries in the diagonal matrices $\Lambda, \hat{\Lambda}$?

c) Show that $S^T CS = G\hat{\Lambda}G^T$ where $G$ is orthogonal.

d) Show that $trace(B - C)^2 = trace((\Lambda - G\hat{\Lambda}G^T)^2)$

e) Show that the quantity $trace((\Lambda - G\hat{\Lambda}G^T)^2)$ is a minimum when $G = I$.

## Problem 5

Show that zero is always an eigenvalue of the Laplacian eigenvector problem. What is the associated eigenvector?

# Part B: Computing

## Problem 1

Use LBG clustering with 10 centers to quantize the color image provided on Canvas. Include the color quantized image in your write-up.

## Problem 2

This is an SOM warm-up problem. Consider the data set consisting of the following 6 points:

| $\mu$ | $x^{(\mu)}$ |
|---|---|
| 1 | 0.34 |
| 2 | 0.12 |
| 3 | 0.73 |
| 4 | 0.97 |
| 5 | 0.07 |
| 6 | 0.56 |

Run SOM on this data set using the index set $\{1, 2, 3, 4, 5, 6\}$ and the Euclidean and Clock metrics. Pick your initial centers randomly from the uniform distribution on the interval $[0, 1]$. Describe what SOM is doing in each case.

## Problem 3

Visualize the animal data in two dimensions using Self-Organizing Mappings.

## Problem 4

Visualize the animal data in two dimensions using Laplacian Eigenmaps.