Homework 1

Part I

Basic Concepts in Linear Algebra and Calculus

1. We have two vectors, x_1 and x_2

$x_1 = \begin{bmatrix}1\\2\end{bmatrix}$ and $x_2 = \begin{bmatrix}10\\18\end{bmatrix}$

What is the distance between x_1 and x_2 ?

(1) if the distance measure is based on L2 norm (a.k.a Euclidean norm)

(2) if the distance measure is based on L1 norm

(3) if the distance measure is based on L∞ norm (a.k.a infinity norm)

Assuming there are two feature components $x = \begin{bmatrix}income\\spend\end{bmatrix}$ in an application, does the L∞ norm-based distance measure make sense for the application of customer segmentation?

2. We define a scalar valued function of a vector variable

$$f(x) = x^T Ax$$

Here, x is a column vector, x^T is the transpose of x, and A is a symmetric matrix

To simplify this question, let's assume x has only two elements $x = \begin{bmatrix}\alpha\\\beta\end{bmatrix}$, and $A = \begin{bmatrix}a & c\\c & b\end{bmatrix}$

The derivative of f with respect to x is a vector defined by $\frac{df}{dx} = \begin{bmatrix}\frac{df}{d\alpha}\\\frac{df}{d\beta}\end{bmatrix}$

Show that $\frac{df}{dx} = 2Ax$

Hint: calculate $f(x)$, $2Ax$, $\frac{df}{d\alpha}$ and $\frac{df}{d\beta}$

K-means clustering

3. Briefly describe the two key steps in one iteration of the k-means algorithm.

4. What is the distance measure used in k-means (implemented in sk-learn)?

5. The k-means algorithm can converge in a finite number of iterations. Why?

6. The clustering result of k-means could be random. Why?

7. The minimum value of the objective/loss function is zero for any dataset. What is the clustering result when the objective function is zero?

Note: for questions 3,4,5,6,7, you only need to write a few words (bullet points) for each one.

You may write the answers on a piece of paper, take a photo using your cell phone, and upload the picture to Blackboard. **Make sure that your handwriting is human-readable**.

You may use MS-word to write the answers, convert the file to PDF, and upload it to Blackboard.

Part 2: Programming

Complete the tasks in the files:

H1P2T1_kmeans.ipynb

If you want to get some bonus points, try this task:

H1P2T2_kmeans_compression.ipynb

Grading: the number of points

	Undergraduate Student	Graduate Student
Basic Concepts in Linear Algebra and Calculus	10	10
K-means clustering	10	10
H1P2T1	30	30
H1P2T2	10 (bonus)	10 (bonus)
Total number of points	50 + 10	50+10

Upload your files (*_your_name.ipynb) to blackboard

Do NOT covert the ipynb files to pdf.