Homework 1 Part I

Basic Concepts in Linear Algebra and Calculus

1. We have two vectors, x_1 and x_2

$x_1 = [^1 2]$ and $x_2 = [^{10} 18]$

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)

$\sqrt{(}\,(1-10)^2 + (2-18)^2$ = 18.3575597507

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

(1+10) + (2 + 18) = 31

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

Max(1, 2) + Max(10, 18) = 20

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?

Since (in most cases) customers cannot spend money they do not have, their income will almost always be higher than their spending, so L∞ norm will just give us the total income of all our customers. This value will never approach 0 and continue to grow the more data we add which does not help us segment customer data at all.

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

$$f(x) = (x^T A) * x$$

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 $x = [\alpha \quad \beta]$, and $A = [a \quad c \quad c \quad b]$

The derivative of f with respect to x is a vector defined by $\frac{df}{dx} = [\mathrm{df}/d\alpha \quad \mathrm{df}/d\beta]$

Show that $df/\mathrm{dx} = 2Ax$
Hint: calculate $f(x)$, $2Ax$, $df/_{\mathrm{d}\alpha}$ and $df/_{\mathrm{d}\beta}$

- $\mathrm{f(x)} = ([\alpha, \beta] * [{}^{\mathrm{a}}{}_{\mathrm{c}} \; {}^{\mathrm{c}}{}_{\mathrm{b}}]) * [{}^{\alpha}{}_{\beta}] = (\mathrm{a}^*\alpha + \mathrm{c}^*\beta, \; \mathrm{c}^*\alpha + \mathrm{b}^*\beta) * [{}^{\alpha}{}_{\beta}] = \alpha(\mathrm{a}^*\alpha + \mathrm{c}^*\beta) + \beta(\mathrm{c}^*\alpha + \mathrm{b}^*\beta) = \alpha\alpha\mathrm{a} + 2\beta\mathrm{c}\alpha + \beta\beta\mathrm{b}$
- $2\mathrm{Ax} = [{}^{2\mathrm{a}}{}_{2\mathrm{c}} \; {}^{2\mathrm{c}}{}_{2\mathrm{b}}] * [{}^{\alpha}{}_{\beta}] = [{}^{2\mathrm{a}\alpha + 2\mathrm{c}\alpha}{}_{2\mathrm{c}\beta + 2\mathrm{b}\beta}]$
- $df/_{\mathrm{d}\alpha} = 2\alpha\mathrm{a} + 2\beta\mathrm{c}$
- $df/_{\mathrm{d}\beta} = 2\mathrm{c}\alpha + 2\beta\mathrm{b}$
- $df/_{dx} = [\mathrm{df}/\mathrm{d}\alpha \; \mathrm{df}/\mathrm{d}\beta] = [{}^{2\mathrm{a}\alpha + 2\mathrm{c}\alpha}{}_{2\mathrm{c}\beta + 2\mathrm{b}\beta}]$

K-means clustering

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

   First, the algorithm finds all data points that are currently closest to each center. Then, the centers are readjusted to the average location of all datapoints it is clustered with.

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

   L2 norm

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

   There is a finite, yet very large, number of potential clustering results, one of which will be some absolute/local minimum loss value. Since the loss never increases and always decreases (as we get closer to the "real" centers with each readjustment) eventually it will stop changing due to there being nothing lower.

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

   Due to either bad random centers being too close to each other, or the data itself may not have obvious clusters with lots of overlapping.

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?

   If a cluster has a loss value of 0, that means the distance to all data points from that center point is 0, so the center point is directly atop all the data points it is clustered with.

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