Machine Learning for Business Analytics
Assignment 2
Due: October/ 24 23:59 (GMT+8)

Instruction

- There are two parts in this assignment. You only have to hand in Part 1 regarding the Julia Programming. **Part 2 will not be graded**; of notice is that similar questions may show up in midterm or final.
- Please hand in your Julia code with a **PDF file**. Wrong format will not be graded.
- **Copying the assignment will result in zero points.**
- Late submission will be graded according to the following rule:
  your grade = original grade × (1 − 0.05*h*), if you submit *h* hours after the deadline.

Part 1 Programming Exercises (100%; 20 % each question)

1. (Chelsea's restaurant) Chelsea is a Michelin chef and he has a restaurant located at the city of Taipei. To run her Michelin restaurant, she needs 100 magic ingredients. For simplicity, her names this 100 ingredients from 1 to 100. Each ingredient corresponds to several properties, including price, quality (freshness) level, and nutrition level. And there are 100 customers in this market, with their location recorded in latitude and longitude on a 2-dimensional grid. The latitude and longitude of the customers are in the following array respectively.

   ```
   Random.seed!(8)
   vcat( [ [0, -0.2] + 0.1*randn(2) for i = 1:30 ],
       [ [0.5, 0.5] + 0.1*randn(2) for i = 1:30 ],
       [ [0.5, -0.5] + 0.1*randn(2) for i = 1:20 ],
       [ [0, 0.4] + 0.06* randn(2) for i = 1:20])
   ```

   Chelsea's restaurant locates at (latitude, longitude) = (0.2, 0). Plot the location of these customers as well as Chelsea's restaurant, mark them with different color.

2. Following question 1, help Chelsea with the following problems.
   a. What is are all distances (in terms of 2-norm) between Chelsea's restaurant and all her customers?
   b. Who is the closest customer to Chelsea's restaurant in terms of 2-norm?
   c. 1-norm in $\mathbf{R}^2$ is calculated as $||x||_1 = |x_1| + |x_2|$ for $x = (x_1, x_2)$, what is the farthest customer to Chelsea's restaurant in terms of 1-norm?
   d. Are the 5 closest customers to Chelsea's restaurant calculated using 1-norm and 2-norm the same?

3. Following question 1, the magic ingredients are as follows: Chelsea wants to use the quality and nutrition level of each ingredients to predict the price of ingredient, following $\widehat{pirce_i} = \beta_{quality} \times quality_i + \beta_{nutrition} \times nutrition_i + v$, The parameters are $\beta_{quality} = 0.5$, $\beta_{nutrition} = 7.1$, and $v = 0.1$. Plot the predicted price on the y-axis and the true price on the x-axis. Which data point is the farthest (in terms of 2-norm) to its prediction?

```
Random.seed!(8)
y = randn(100)
Random.seed!(8)
quality = y * 0.5 + rand(100) * 0.1
Random.seed!(10)
nutrition =  y * 0.1 + rand(100) * 0.2
price = quality * 1 + nutrition * 0.1 + rand(100) * 2
```

4. Following question 1, Chelsea is retiring, three of her friends, Maximilian, Jemima, and Say-zed, wants to be join the franchises. Each of their new restaurant will be at the center of k-means algorithm with three being their clustering center. Segment the existing customers according to their geographic location into three groups using the k-means algorithm.
   a. What is best location to open these three restaurants for Chelsea's franchisee, i.e., where are the three clustering centres?
   b. Mark these customers group with different colors and mark these new restaurants on the map.
   c. Use 1-norm to measure distance in the k-means algorithm. And plot the customers on the map. Are they the same as (b)?
   d. Do you think 3 is a good clustering number? Why?

5. (Ida the mathematician) Write a function Gram_Schmidt that takes the input of an array containing vectors. The function returns the orthonormal set of vectors when the input array contains vectors that are all linearly independent; otherwise, return "Vectors are linearly dependent.".

Part 2 Mathematical Exercises (0%)
Chapter 3 Exercises

1. Can we generalize the triangle inequality to three vectors? That is, do we have
$$\|a + b + c\| \leq \|a\| + \|b\| + \|c\|$$
for any *n*-vectors *a*, *b*, *c*? If so, justify it. If it's not true, give a counter-example, i.e., specific vectors *a*, *b*, and *c* for which the inequality is false.

2. Which of the following are true? (True means it holds for any vector x)
   (a) $\text{std}(x) \leq \text{rms}(x)$
   (b) $\text{rms}(x) \leq |\text{avg}(x)|$
   (c) $\text{std}(x)^2 = \text{rms}(x)^2 - \text{avg}(x)^2$
   (d) $\text{std}(x) = 0$ only when all entries of x are equal

3. Let *x* and *y* be (nonzero) word count vectors for two documents, associated with a given dictionary. Choose one in each of the subproblems below, and briefly justify your answer.
   (a) If the angle between *x* and *y* is 0, we can conclude
      (1) The documents are identical
      (2) The documents share no dictionary words
      (3) The word count histogram vectors of the two documents are the same
      (4) The documents have the same author
      (5) The documents have the different authors
   (b) If the angle between *x* and *y* is 90 degree, we can conclude
      (1) The documents are identical
      (2) The documents share no dictionary words
      (3) The word count histogram vectors of the two documents are the same
      (4) The documents have the same author
      (5) The documents have the different authors

4. Suppose x is a 20-vector with $\|x\| = 10$. Which of the statements below follow from the Chebyshev inequality? (We do not count statements that are true, but don't follow from the Chebyshev inequality.)
      (1) No x_i can satisfy $|x_i| \geq 10.5$
      (2) At least one x_i must have magnitude at least 2
      (3) No more than 3 entries of x can exceed 6 in magnitude
      (4) No more than half the entries of x are positive

5. Show that $\angle(x,y) \leq \angle(x,z) + \angle(z,y)$ for any nonzero vectors *x, y, z*. In other words, angles satisfy the triangle inequality. (Recall that angles are normalized to lie between 0 and π)

Chapter 4 Exercises

1. Which of the following statements are correct?
      (a) The goal of clustering a set of vectors is to choose the best vectors from the set
      (b) The goal of clustering a set of vectors is to divide them into groups of vectors

that are near each other

(c) The goal of clustering a set of vectors is to determine the nearest neighbors of each of the vectors

(d) The *k*-means algorithm always converges to a clustering that minimizes the mean-square vector-representative distance

(e) The *k* -means algorithm can converge to different final clusterings, depending on the initial choice of representatives

(f) The *k* -means algorithm is widely used in practice

(g) The choice of *k*, the number of clusters to partition a set of vectors into, depends on why you are clustering the vectors

(h) The choice of *k*, the number of clusters to partition a set of vectors into, should always be as large as your computer system can handle

Chapter 5 Exercises

1. Suppose a_1, a_2 is a list of two linearly independent n-vectors. When we run the Gram-Schmidt algorithm on this list, we obtain the orthonormal vectors q_1, q_2. Now suppose we run the Gram-Schmidt algorithm on the list of vectors a_2, a_1 (i.e., the same vectors, in reverse order). Do we get the orthonormal vectors q_1, q_2 (i.e., the orthonormal vectors obtained from the original list, in reverse order)? If you believe this is true, give a very brief explanation why. If you believe it is not true, give a simple counter-example.

Chapter 6 Exercises

1. Suppose $A$is an $n \times n$ matrix. When is $A + A^T$ symmetric? You can answer never, sometimes, or always. Justify your answer.

2. Suppose $Ax = 0$, where A is an $m \times n$ matrix and x is a nonzero *n*-vector. Then we can conclude
   (1) $A = 0$
   (2) The rows of A are linearly dependent
   (3) The columns of A are linearly dependent

3. The n-vector p gives the daily time series of the price of an asset over n trading days, with $n \geq 4$. The $(n - 3)$-vector d gives the difference of the
   current asset price and the average asset price over the previous three trading days, starting from the fourth day. Specifically, for $i = 1, \dots, n - 3$, we have $d_i = p_{i+3} - (p_i + p_{i+1} + p_{i+2})/3$. (Note that d is an $(n - 3)$-vector.)
   Given the matrix *A* for which $d = Ap$, for the specific case $n = 6$. Be sure to give its size all entries.