Homework 2

Submission Instructions

- Homework is due on: **Thursday 15/08/19 23:55**.
- Homework should be done **only in pairs**. Each pair is to do their own work, separate from the other pairs.
- We prefer you type your submission, however, you may submit scanned handwritten material as long as it is **clear and readable**.
- Submit **only one** PDF file. Please **write your ID** on the top of the file.
- Submission is done via **Moodle** website.
- Homework can be done using either MATLAB or Python.

Question 1

Vectors $x \in \mathbb{R}^6$ are generated in the following manner:

First, we sample $z \sim \mathcal{N}(\mu, \Sigma)$ where

$$\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 21 & 0 & 0 \\ 0 & 0 & 13 & 8 \\ 0 & 0 & 8 & 13 \end{bmatrix}.$$

Then, apply a linear transform $A \in \mathbb{R}^{6\times 4}$ on z to get $x = Az$ where $A^T A = I$ (See Tutorial 6).

(a) Calculate the eigenvalues of the covariance matrix Σ_x of the random vector x. Compute the sum of eigenvalues denoted by S_λ.

(b) How many PCA directions should be taken so that the mean reconstruction error will be smaller than $S_\lambda/4$?

(c) We define new vector $y \triangleq x + v \in \mathbb{R}^6$ where $v \in \mathbb{R}^6$. The vector $v \sim \mathcal{N}(0, \Sigma_z)$ is normally distributed where

$$\Sigma_z = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix}.$$

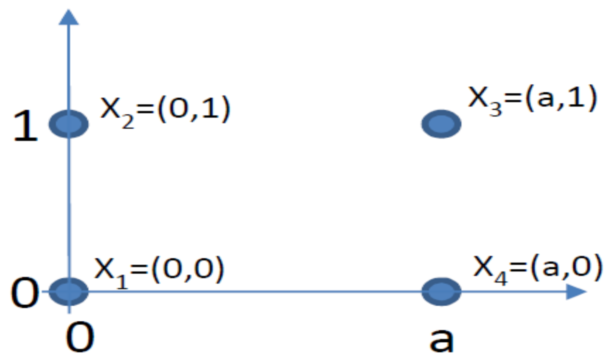
Repeat sections (a) and (b) for the vector y (note that the sum of the eigenvalues has changed).

Question 2

Reminder: The K-Means algorithm divides samples into clusters, and changes a sample's cluster association from cluster A to cluster B only if its distance to the centroid of group B is strictly smaller than its distance to the centroid of group A.

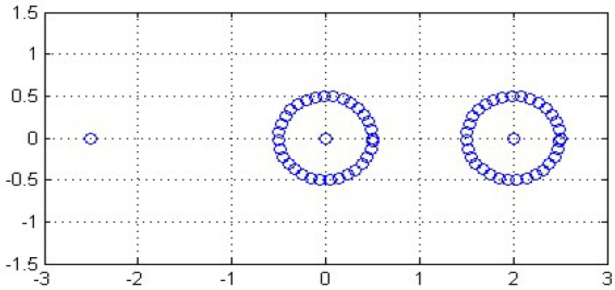
Fix-Point: A specific clustering of the samples into certain clusters is called a **fix point** of K-Means, if the algorithm does not change the sample clustering when it used as initial conditions.

Consider a training set of four samples in $\mathbb{R}^2$ located on the vertices of a rectangle parallel to the axes (see figure below). Find all possible fixed point divisions/clusterings of the four samples into two groups. Your answer may depend on the rectangle's width parameter $a > 0$.



Question 3

For the set of samples shown in the figure below and for K-Means with k=2, is there one option for convergence or more? If only one option exists, write it down and explain why it is the only option. If there is more than one option, mention at least two other options, and explain which one of them attains a lower value of the objective function.



Question 4

Reminder: The perceptron learning algorithm

Input: set of labeled examples $\{x_i, d_i\}_{i=1}^{n}$ where $d_i \in -1, 1$ and $x_i \in \mathbb{R}^m$.
Initialization: the weights vector w_0 is initialized with zeros.
For each step $t = 1, 2, ..$:

- Choose one example x_t from the dataset.
- Calculate the perceptron output for that sample using the current weight vector w_t:

$$y_t = sign(w_t^T x_t).$$

- Update the weights vector - $w_{t+1} = w_t + \eta(d_t - y_t)x_t$.

Assume

- $sign(0) = 0$.
- $\eta = \frac{1}{2}$.
- The example x_t that the algorithm receives at step t is one of the columns of an orthogonal matrix with size $D \times n$ (i.e. the norm of each example is one $||x_t||_2 = 1$).
- The dataset is linearly separable.

How many updates will the algorithm preform, at the least?

Solution

Question 1

(a) The eigenvalues of the non-diagonal sub-matrix of Σ are given by the solutions of $(13-\lambda)^2 - 64 = 0$, which are $\lambda = 21, 5$. The eigenvalues of the covariance matrix of z are therefore $3, 21, 21, 5$. The covariance matrix of x is given by

$$\Sigma_x = A\Sigma A^T,$$

hence, the eigenvalues of the covariance matrix of x are $3, 21, 21, 5, 0, 0$ and their sum is 50. Another way to see it is

$$Tr(\Sigma_x) = Tr(A\Sigma A^T) = Tr(A^T A\Sigma) = Tr(\Sigma) = 50.$$

(b) The reconstruction error is the sum of the eigenvalues discarded by the PCA. Denote by m the number of directions of PCA. We want that the sum of the m largest eigenvalues is at least $0.75 * 50 = 37.5$. Therefore two directions are enough, because $21 + 21 > 37.5$.

(c) Notice that $\Sigma_y = \Sigma_x + \Sigma_v = \Sigma_x + 5I$, hence, We the value 5 is added to each of the eigenvalues, and we get the eigenvalues are $5, 5, 10, 26, 26, 8$ and their sum is 80. We want the sum of the m largest eigenvalues is at least $0.75 * 80 = 60$. Therefore, three directions are enough - $26 + 26 + 10 > 60$.

Question 2

All solutions which cluster the 4 samples to 2 groups of size 2 are possible:

(a) $\{1, 2\}, \{3, 4\}$ - centers are $(0, 0.5)$ and $(a, 0.5)$.

(b) $\{1, 4\}, \{2, 3\}$ - centers are $(0.5a, 0)$ and $(0.5a, 1)$.

(c) $\{1, 3\}, \{2, 4\}$ - centers are identical $(0.5a, 0.5)$, the solution is stable because for instability we need strict inequality.

A solution where one cluster includes all samples and the other cluster is empty is also possible, by initializing one of the centers to be far enough of all samples.

We now check solutions with 3 samples in one cluster and 1 sample in the other. From symmetry it is enough to check one such solution, for example $\{1, 2, 3\}, \{4\}$. Centers are $(a/3, 2/3)$ and $(a, 0)$. It is clear that x_4 and x_2 each are closer to the center of their cluster. What about x_1 and x_3? A condition for stability for x_1 is

$$\begin{aligned}
&||(0,0) - (1/3a, 2/3)||^2 \leq ||(a,0) - (0,0)||^2 \\
&\Rightarrow 4/9 + 1/9a^2 \leq a^2 \\
&\Rightarrow 0.5 \leq a^2.
\end{aligned}$$

A condition for stability for x_3 is

$$\begin{aligned}
&||(a,1) - (1/3a, 2/3)||^2 \leq ||(a,1) - (a,0)||^2 \\
&\Rightarrow 4/9a^2 + 1/9 \leq 1 \\
&\Rightarrow a^2 \leq 2.
\end{aligned}$$

So we conclude that for $\frac{1}{\sqrt{2}} \leq a \leq \sqrt{2}$ we have 4 more stable clusterings.

Question 3

There may possible options for convergence with $K = 2$, for example:

- $\{(0 - \Delta,), (2, 0)\}$
- $\{(1, 0), (-2.5, 0)\}$
- $\{(1 - \Delta, 0), (10, 10)\}$

where Δ is a small constant. The global optimum of the K-Means objective function is attained when the centers are at the center of the 2 main clusters, that is $\{(0 - \Delta,), (2, 0)\}$.

Question 4

The minimal number of update is n. The optimal solution is of the form $w = \sum_i \alpha_i x_i$. Since we assume $sign(0) = 0$ and $x_i^T x_j = 0$ $(i \neq j)$ we get

$$d_i = sign(w^T x_i) = sign(\alpha_i),$$

hence, all the coefficients α_i should be non-zeros. Considering that each update sets one coefficient α to be non-zero at most, the algorithm must perform n updates at least.