Prof. Dr. Rudolf Mathar, Dr. Gholamreza Alirezaei, Dr. Arash Behboodi

Exercise 9
- Proposed Solution -

Friday, January 20, 2017

Solution of Problem 1

a) (2-Classes) For the dataset with two classes, Figure 1 presents the results of K-means clustering algorithm.

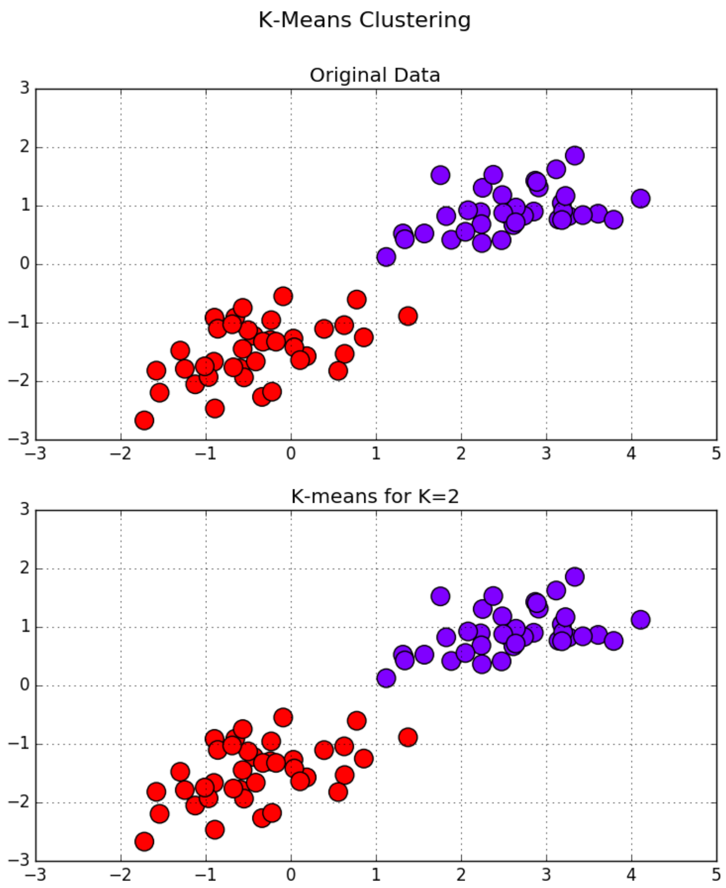

Figure 1: Dataset with two classes

b) (3-Classes) For the dataset with three classes, Figure 2 presents the results of K-means clustering algorithm.

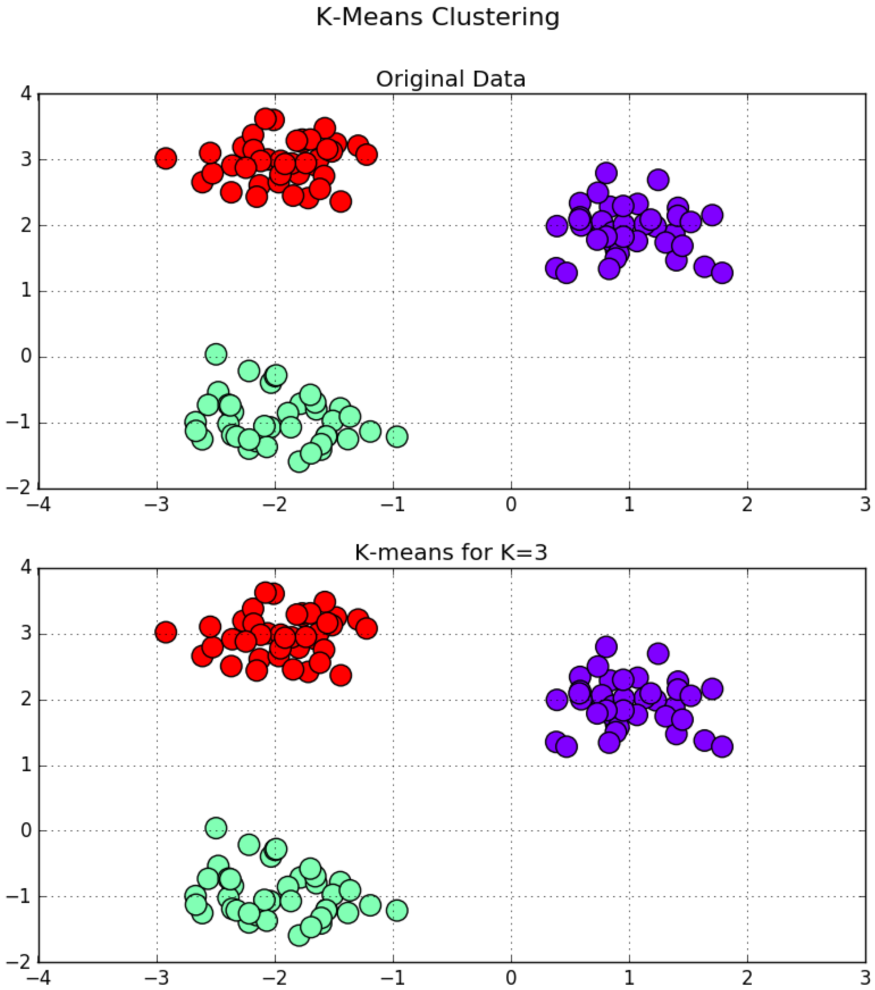

Figure 2: Dataset with three classes

Solution of Problem 2

a) (2-Classes) For the dataset with two classes, Figure 1 presents the results of K-means clustering algorithm. Note that the circled points are wrongly clustered points compared to the original labels.

b) (3-Classes) For the dataset with three classes, Figure 2 presents the results of K-means clustering algorithm. Note that the circled points are wrongly clustered points compared to the original labels.

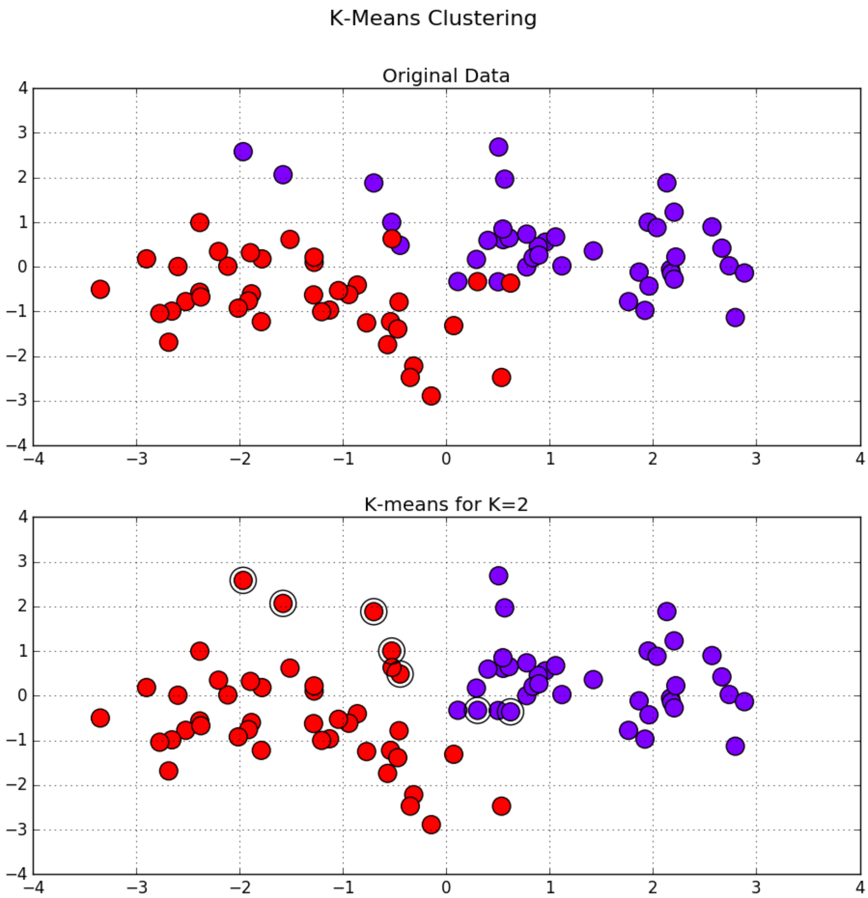

Figure 3: Dataset with two classes

Solution of Problem 3

a) The first step is to formulate the Lagrangian function:

$$\Lambda(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{c}^T\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = (\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda})^T\mathbf{x} - \boldsymbol{\lambda}^T\mathbf{b}$$

$\boldsymbol{\lambda}$ is Lagrange multiplier. Next step is to find Lagrange dual function:

$$g(\boldsymbol{\lambda}) = \inf_{\mathbf{x}\in\mathbb{R}^n} \Lambda(\mathbf{x}, \boldsymbol{\lambda}) = \begin{cases} -\boldsymbol{\lambda}^T\mathbf{b} & \mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda} = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Therefore the dual problem can be written as:

$$\begin{aligned} \max \quad & -\mathbf{b}^T\lambda \\ \text{s.t.} \quad & \mathbf{A}^T\boldsymbol{\lambda} + \mathbf{c} = 0 \\ & \boldsymbol{\lambda} \succeq 0 \end{aligned}$$

b) Again, the first step is to formulate the Lagrangian function:

$$\Lambda(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{x}^T\mathbf{B}\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

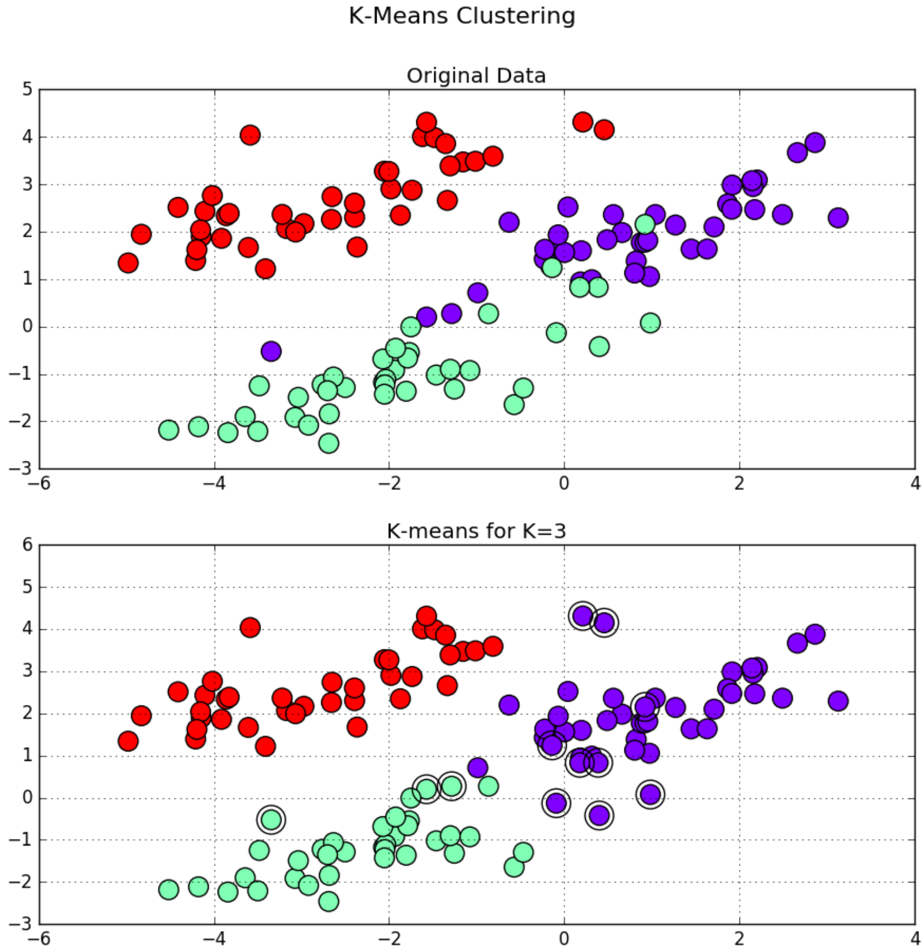

Figure 4: Dataset with three classes

Next step is to find Lagrange dual function:

$$g(\boldsymbol{\lambda}) = \inf_{\mathbf{x}\in\mathbb{R}^n} \Lambda(\mathbf{x}, \boldsymbol{\lambda})$$

Using derivation with repsect to $\mathbf{x}$, we have:

$$\frac{\partial}{\partial \mathbf{x}}\Lambda(\mathbf{x}, \boldsymbol{\lambda}) = 2\mathbf{B}x + \mathbf{A}^T\boldsymbol{\lambda} \implies \mathbf{x} = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{A}^T\boldsymbol{\lambda}.$$

Having this solution (verify that it is the minimum point indeed), the Lagrange dual function is given by:

$$g(\boldsymbol{\lambda}) = \frac{1}{4}\boldsymbol{\lambda}^T\mathbf{A}\mathbf{B}^{-1}\mathbf{A}^T\boldsymbol{\lambda} + \boldsymbol{\lambda}^T(-\frac{1}{2}\mathbf{A}\mathbf{B}^{-1}\mathbf{A}^T\boldsymbol{\lambda} - \mathbf{b}) = -\frac{1}{4}\boldsymbol{\lambda}^T\mathbf{A}\mathbf{B}^{-1}\mathbf{A}^T\boldsymbol{\lambda} - \boldsymbol{\lambda}^T\mathbf{b}$$

Therefore the dual problem can be written as:

$$\begin{aligned} \max \quad & -\frac{1}{4}\boldsymbol{\lambda}^T\mathbf{A}\mathbf{B}^{-1}\mathbf{A}^T\boldsymbol{\lambda} - \boldsymbol{\lambda}^T\mathbf{b} \\ \text{s.t.} \quad & \boldsymbol{\lambda} \succeq 0 \end{aligned}$$

c) Again, the first step is to formulate the Lagrangian function:

$$\Lambda(\mathbf{x}, \boldsymbol{\nu}) = \|\mathbf{x}\|_p + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

Next step is to find Lagrange dual function:

$$g(\boldsymbol{\nu}) = \inf_{\mathbf{x}\in\mathbb{R}^n} \Lambda(\mathbf{x}, \boldsymbol{\nu})$$

A related optimization problem can be written as:

$$\inf_{\mathbf{x}} \|\mathbf{x}\|_p + \mathbf{a}^T\mathbf{x}.$$

Consider the Hölder's inequality:

$$\|\mathbf{x}\|_p\|\mathbf{a}\|_q \geq |\mathbf{a}^T\mathbf{x}| \geq -\mathbf{a}^T\mathbf{x},$$

if $\frac{1}{p}+\frac{1}{q} = 1$. Note that if $\|\mathbf{a}\|_q \leq 1$, then $\|\mathbf{x}\|_p+\mathbf{a}^T\mathbf{x} \geq 0$ and therefore $\inf_{\mathbf{x}} \|\mathbf{x}\|_p+\mathbf{a}^T\mathbf{x} = 0$.

Now suppose that $\|\mathbf{a}\|_q > 1$. See that if $\mathbf{x} = -t(a_1|a_1|^{q-2}, \ldots, a_n|a_n|^{q-2})$ for some $t > 0$, then we have:

$$\|\mathbf{x}\|_p + \mathbf{a}^T\mathbf{x} = t\|\mathbf{a}\|_q^{q-1} - t\|\mathbf{a}\|_q^q = t(\|\mathbf{a}\|_q^{q-1} - \|\mathbf{a}\|_q^q).$$

Since $\|\mathbf{a}\|_q > 1$, the above expression is strictly negative and can be arbitrarily small when $t \to \infty$.

Applying the above result, the Lagrange dual function is given by:

$$g(\boldsymbol{\nu}) = \inf_{\mathbf{x}\in\mathbb{R}^n} \Lambda(\mathbf{x}, \boldsymbol{\nu}) = \begin{cases} -\boldsymbol{\nu}^T\mathbf{b} & \|\mathbf{A}^T\boldsymbol{\nu}\|_q \leq 1 \\ -\infty & \text{otherwise.} \end{cases}$$

Therefore the dual problem can be written as:

$$\begin{aligned} \max \quad & -\boldsymbol{\nu}^T\mathbf{b} \\ \text{s.t.} \quad & \|\mathbf{A}^T\boldsymbol{\nu}\|_q \leq 1 \end{aligned}$$