CSCI 567: Homework 5

Deepika Anand

November 10, 2016

Problem 1. (a)

To prove that D is minimized μ_k is the mean, differentiate D with respect to μ_k and equate it to 0

$$\frac{\partial D}{\partial \mu_k} = \frac{\partial}{\partial \mu_k} \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} ||x_n - \mu_k||^2 \tag{1}$$

$$\frac{\partial D}{\partial \mu_k} = \frac{\partial}{\partial \mu_k} \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} (x_n - \mu_k)^T (x_n - \mu_k) \tag{2}$$

$$\frac{\partial D}{\partial \mu_k} = \frac{\partial}{\partial \mu_k} \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} (x_n^T - \mu_k^T)(x_n - \mu_k) \tag{3}$$

$$\frac{\partial D}{\partial \mu_k} = \frac{\partial}{\partial \mu_k} \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} (x_n^T x_n - x_n^T \mu_k - x_n \mu_k^T + \mu_k^T \mu_k) \tag{4}$$

$$\frac{\partial D}{\partial \mu_k} = \sum_{n=1}^{N} r_{nk} (2\mu_k - 2x_n) \tag{5}$$

Equating the equation to 0

$$\mu_k = \frac{\sum_{n=1}^{N} r_{nk} x_n}{\sum_{n=1}^{N} r_{nk}} \tag{6}$$

Hence D is minimized when μ_k is the mean of points

Problem 1. (b)

Distortion measure D can be written as

$$D = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} sign(x_n - \mu_k) \tag{7}$$

where

$$sign(x_n - \mu_k) = +1 if x_n - \mu_k > 0 \tag{8}$$

$$sign(x_n - \mu_k) = -1 if x_n - \mu_k < 0 \tag{9}$$

Hence,

$$\sum_{n=1}^{N} sign(x_n - \mu_k) = 0 \tag{10}$$

There will be a point x_n such that Number of points, given by function Num such that Num$(x_n|x_n - \mu_k > 0)$ - Num$(x_n|x_n - \mu_k < 0) = 0$
That point will x_n be median

Problem 1. (c) (a)

$$\tilde{D} = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}||\phi(x_n) - \tilde{\mu_k}||^2 \tag{11}$$

where

$$\tilde{\mu_k} = \frac{\sum_{i=1}^{N}\phi(x_i)}{\sum_{n=1}^{N} r_{nk}} \tag{12}$$

$$\tilde{D} = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}(\phi(x_n) - \frac{\sum_{i=1}^{N}\phi(x_i)}{\sum_{n=1}^{N} r_{nk}})^T(x_n - \frac{\sum_{i=1}^{N}\phi(x_i)}{\sum_{n=1}^{N} r_{nk}}) \tag{13}$$

$$\tilde{D} = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}(\phi(x_n)^T\phi(x_n) - \frac{2\sum_{i=1}^{N}\phi(x_n)^T\phi(x_i)}{\sum_{n=1}^{N} r_{nk}} + \sum_{i=1}^{N}\frac{\phi(x_i)^T\phi(x_i)}{(\sum_{n=1}^{N} r_{nk})^2}) \tag{14}$$

$$\tilde{D} = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}(k(x_n, x_n)) - \frac{2\sum_{i=1}^{N} k(x_i, x_n)}{\sum_{n=1}^{N} r_{nk}} + \sum_{i=1}^{N}\frac{k(x_i, x_i)}{(\sum_{n=1}^{N} r_{nk})^2}) \tag{15}$$

Hence, we can show $\tilde{D}$ can be represented only in the form of $k(x_i, x_j) = \phi(x_i)^T\phi(x_j)$

Problem 1. (c) (b)

When we are given a data point x_n we select arg j such that cluster $k =$ arg min $(||x_n - \mu_j||)_2^2$.
Replacing $\mu_j = \frac{\sum_{x_n \in C_j}\phi(x_n)}{|x_n \in C_j|}$

$$= argmin_j[\phi(x)^T\phi(x) - \frac{2\sum_{x_n \in C_j}\phi(x_n)^T\phi(x)}{|x_n \in C_j|} + \frac{\sum_{x_n \in C_j}\phi(x_n)^T\phi(x_n)}{|x_n \in C_j|^2}] \tag{16}$$

Replacing with $k(x_i, x_j) = \phi(x_i)^T\phi(x_j)$

$$k = argmin_j[k(x, x) - \frac{2\sum_{x_n \in C_j} k(x_n, x)}{|x_n \in C_j|} + \frac{\sum_{x_n \in C_j} k(x_n, x_n)}{|x_n \in C_j|^2}] \tag{17}$$

Hence,

$$\mu_k = argmin_j[k(x, x) - \frac{2\sum_{x_n \in C_j} k(x_n, x)}{|x_n \in C_j|} + \frac{\sum_{x_n \in C_j} k(x_n, x_n)}{|x_n \in C_j|^2}] \tag{18}$$

Problem 1. (c) (c)

Pseudo code for kernel k-means

```
1. C_{1}, C_{2},....., C_{k} <- Assign random centers
2. for x_{n} \n Input data
    for c in [C_{1}, C_{2},....., C_{k}]
        2.1 Compute ||\phi(x_{n} - \mu_{k}) distance using equation (11)
    Assign cluster with minimum distance, k to C*(x_{n})
3. If converged:
    return Clusters
   else:
    Go back to Step 2
```

Problem 2.

Putting, $\mu_1 = 0$ and $\sigma^2 = 1$ in Gaussian equation for $p(x_1|\theta = \theta_1)$

$$p(x = x_1|\theta = \theta_1) = \frac{e^{-x_1/2}}{\sqrt{2 * \pi}} \tag{19}$$

Putting, $\mu_1 = 0$ and $\sigma^2 = 0.5$ in Gaussian equation for $p(x = x_1|\theta = \theta_2)$

$$p(x = x_1|\theta = \theta_2) = \frac{e^{-x_1/0.25}}{\sqrt{2 * \pi}} \tag{20}$$

$$p(x) = (\alpha) * p(x = x_1|\theta = \theta_1) + (1 - \alpha) * (p(x = x_1|\theta = \theta_2)) \tag{21}$$

Likelihood will be given as ,

$$logp(x) = log\alpha - log\sqrt{2 * \pi} - \frac{x_1}{2} + log(1 - \alpha) - log\sqrt{2 * \pi} - \frac{x_1}{0.25} \tag{22}$$

Now differentiating the above equation wrt α, we get

$$= \frac{1}{\alpha} - \frac{1}{1 - \alpha} \tag{23}$$

Equating the equation equal to 0, we get $\alpha = 0.5$

Problem 3. (a)

Given the p(x_i). Hidden variable z_i will be such that $z_i = 0$, if x_i is from Poisson distribution else $z_i = 1$
$L(\pi, \lambda)$ will be given as

$$= \prod_{x_i=0} \pi^{z_i} * ((1 - \pi)e^{-\lambda})^{1-z_i} * \prod_{x_i>0} (1 - \pi)^{\frac{\lambda_i^x * e^{-\lambda}}{x_i!}} \tag{24}$$

Taking Log on both sides

$$LogL(\pi, \lambda) = \sum_{x_i=0} z_i log\pi + (1 - z_i)(log(1 - \pi) - \lambda) + \sum_{x_i>0} log(1 - \pi) + x_i log\lambda - log(x_i!) - \lambda \tag{25}$$

Problem 3. (b)

E-step is finding Q such that $Q = E[L(\pi, \lambda)]$
E-Step

$$Q(\theta, \theta_0) = \sum_{x_i=0} E_{p(z|x)} z_i log\pi + (1 - E_{p(z|x)} z_i)(log(1-\pi) - \lambda) + \sum_{x_i>0} log(1-\pi) + x_i log\lambda - log(x_i!) - \lambda \tag{26}$$

$$E_{p(z|x)} = \frac{p(x_i|z_i = 0) * p(z_i = 0)}{p(x_i|z_i = 0) * p(z_i = 0) + p(x_i|z_i = 1) * p(z_i = 1)} \tag{27}$$

$$E_{p(z|x)} = \frac{\pi_0}{\pi_0 + (1 - \pi_0)e^{-\lambda_0}} \tag{28}$$

M-Step is Maximization of Q
Hence differentiate with respect λ and π
Find $\frac{\partial Q}{\partial \lambda}$ and equate it to 0

$$= \sum_{x_i=0} (1 - E_{p(z|x) z_i})(-1) + \sum_{x_i>0} (\frac{x_i}{\lambda} - 1) \tag{29}$$

$$\hat{\lambda} = \frac{\sum_{x_i>0} x_i}{n - \sum_{x_i=0} E[z_i]} \tag{30}$$

where,

$$E[z_i] = \frac{\pi_0}{\pi_0 + (1 - \pi_0)e^{-\lambda_0}} \tag{31}$$

Find $\frac{\partial Q}{\partial \pi}$ and equate it to 0

$$= \sum_{x_i=0} \frac{E[z_i]}{\pi} - \frac{1 - E[z_i]}{1 - \pi} - \frac{n}{1 - \pi} \tag{32}$$

$$\hat{\pi} = \sum_{x_i=0} \frac{E[z_i]}{n} \tag{33}$$

where,

$$E[z_i] = \frac{\pi_0}{\pi_0 + (1 - \pi_0)e^{-\lambda_0}} \tag{34}$$

Problem 4. .1

Imported data : Circle.csv had shape of (500, 2)
Blob.csv had shape of (600, 2)

Problem 4. .2

k-means fails to separate the two circles in *circle.csv* because the assumption was made that data is linearly separable however that was not the case. so we need to map data set to higher dimension so that they can be separated using hyperplane.

Problem 4. .3

Kernel

$$K(x_i, x_j) = \sqrt{\sum x_i^2} * \sqrt{\sum x_j^2} \tag{35}$$

Problem 4. .4

μ_1 = (-0.32536411, 0.97109313)
μ_2 = (0.75896585, 0.67976677)
μ_3 = (-0.63954432, 1.47408141)
σ_1 =

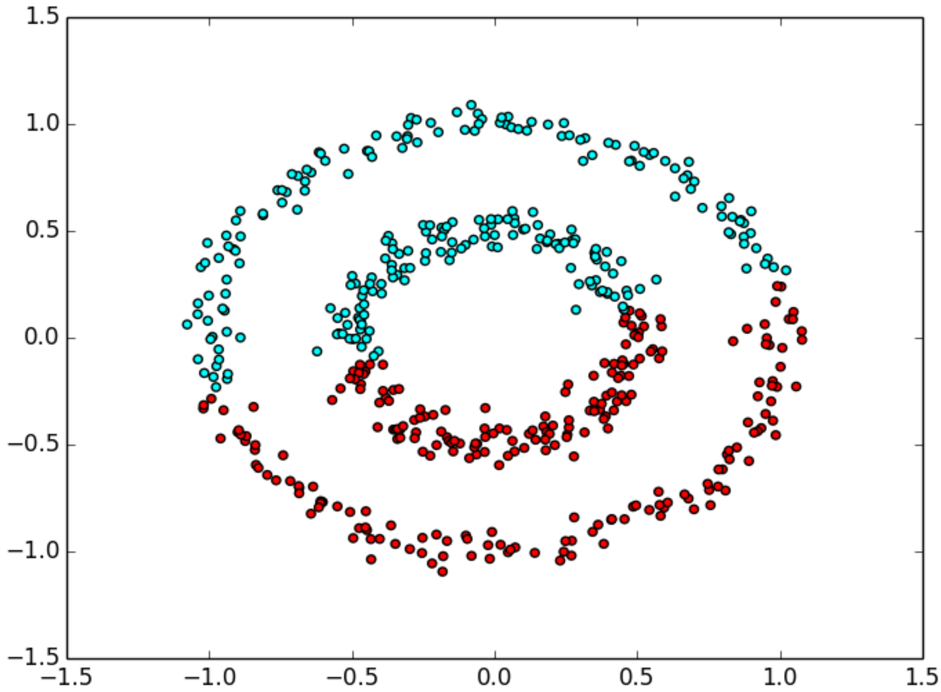
```
 0.03609241,  0.01482509
 0.01482509,  0.01635765
```

σ_2 =

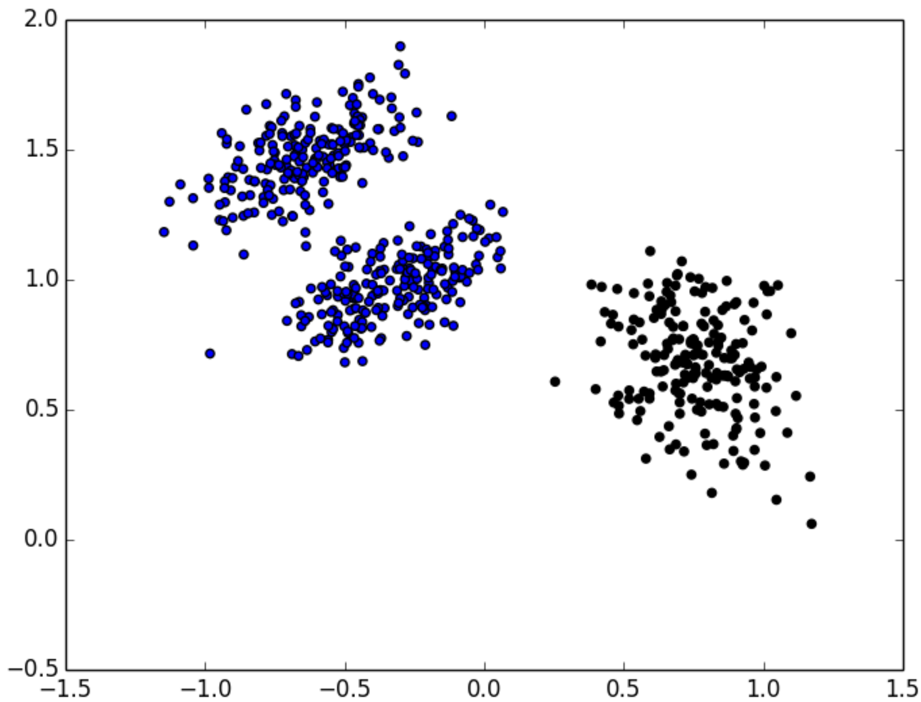
```
 0.02730268, -0.00844004
-0.00844004,  0.04064383
```

σ_3 =

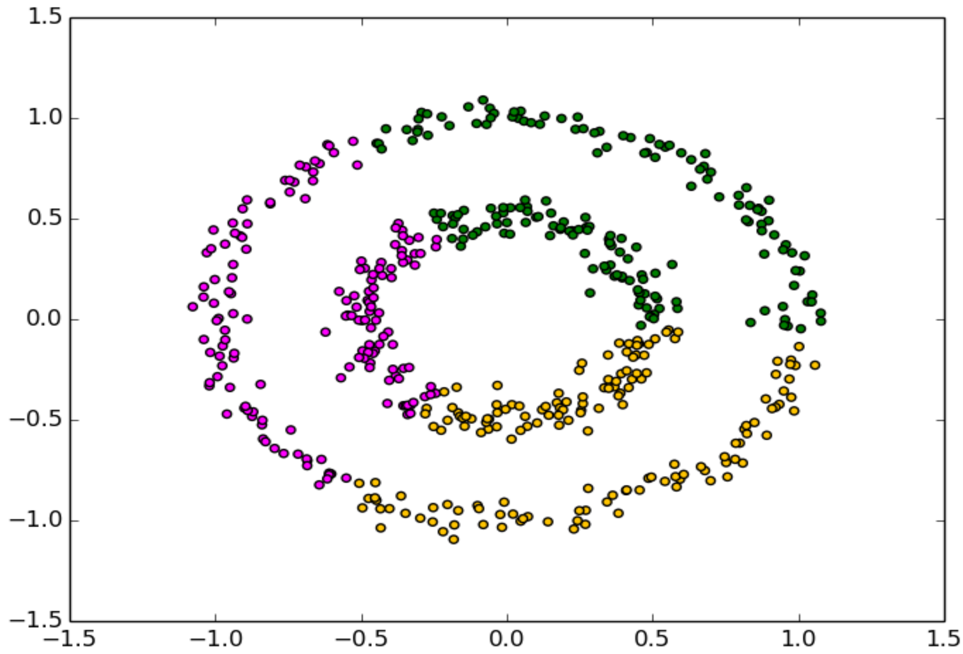
```
 0.0360839 ,  0.01557532
 0.01557532,  0.0196006
```



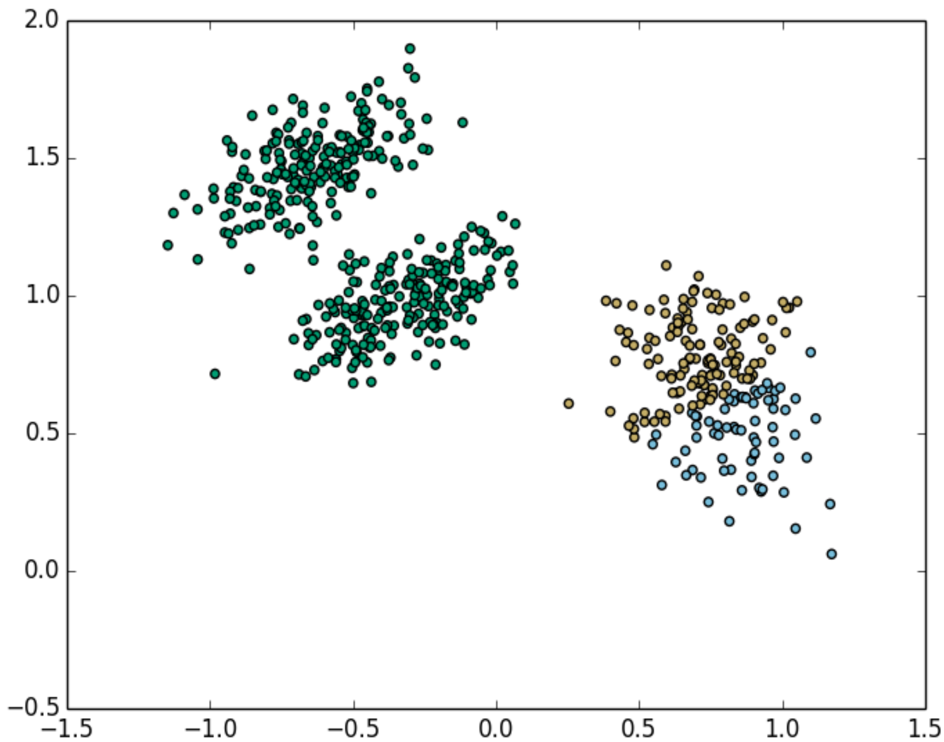
k=2(Circle Data set)



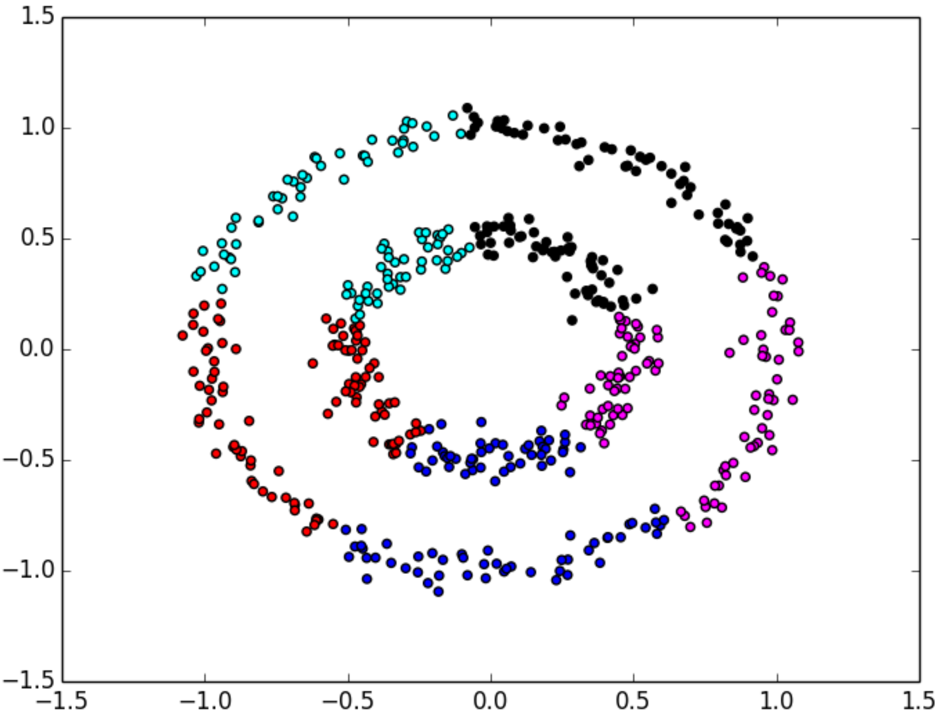
k=2(Blob Data set)



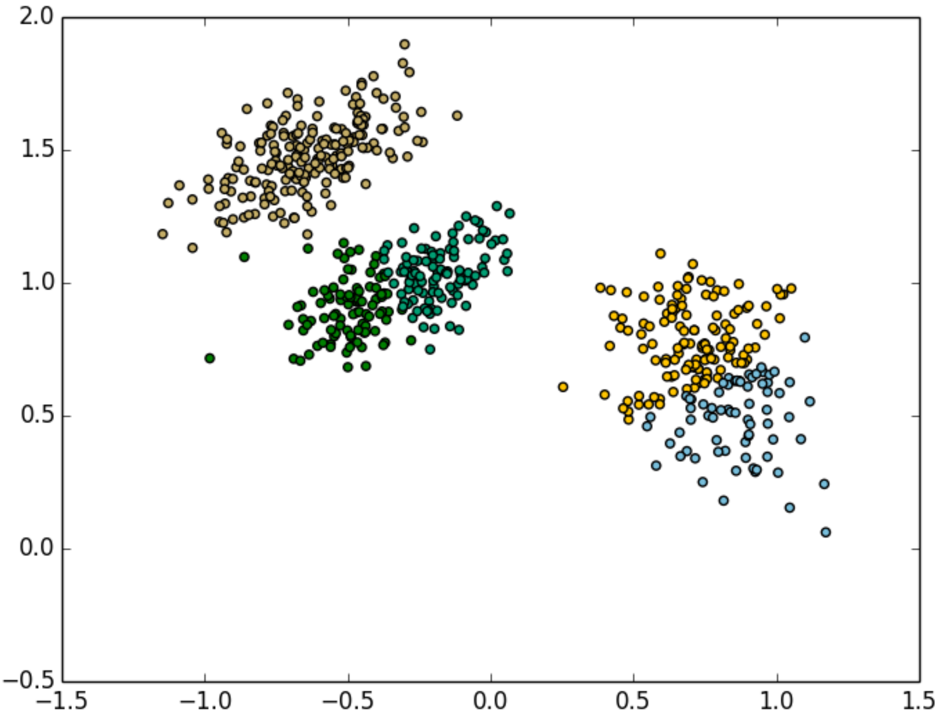
k=3 (Circle Data set)



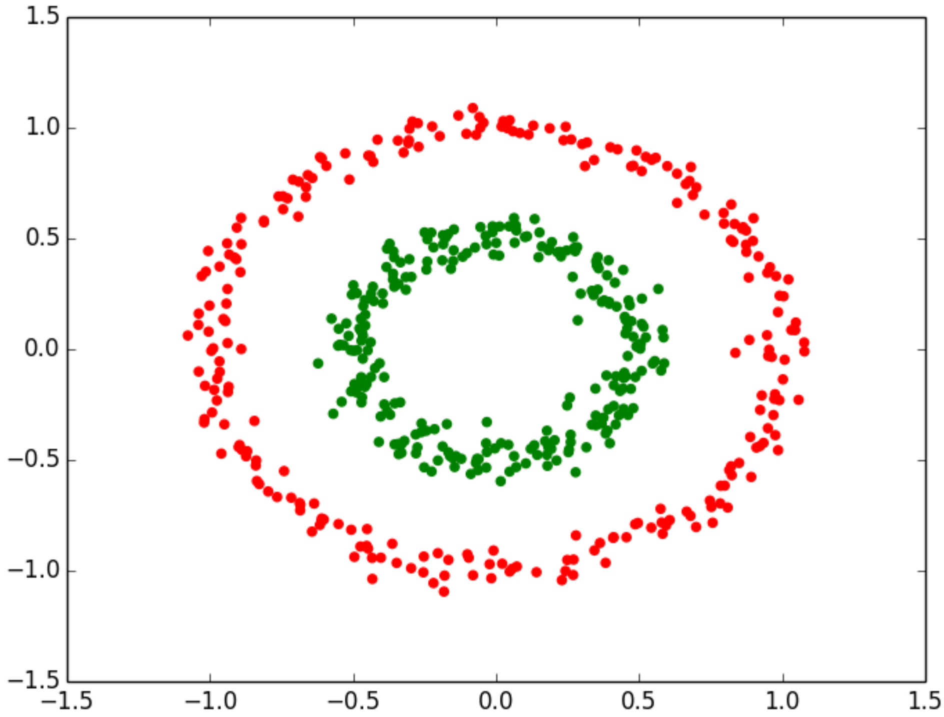
k=3 (Blob Data set)



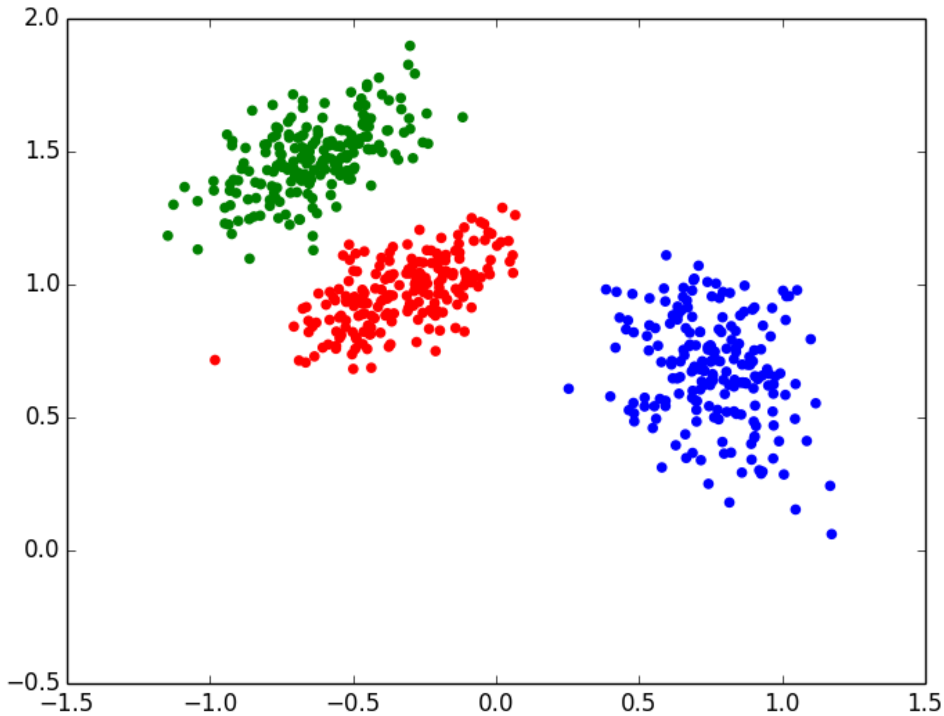
k=5(Circle Data Set)



k=5 (Blob data set)



Kernel K-Means k=2 (Circle data set)



Gaussian Mixture Model