CSCI-567: Assignment #5

Due on Tueday, November 17, 2015(One late day used)

Saket Choudhary
2170058637

Contents

Problem 1

Problem 1:   (a)

To find $\nabla_{y_t} L$:

$$\nabla_{y_t} L = \frac{\partial}{\partial y_t} \frac{1}{2} \sum_{i=1}^{N} (y_i - \hat{y}_i)^T (y_i - \hat{y}_i)$$

$$= \frac{\partial}{\partial y_t} \frac{1}{2} \sum_{i=1}^{N} (y_i^T y_i - 2 y_i^T \hat{y}_i + \hat{y}_t^2)$$

$$= \frac{1}{2} (2 y_t - 2 \hat{y}_t)$$

$$= y_t - \hat{y}_t$$

$$\boxed{\nabla_{y_t} L = y_t - \hat{y}_t}$$

Problem 1: (b)

To find $\nabla_{y_t} L$:

$$\nabla_{s_t} L = \sum_{k=1}^{T} \frac{\partial L}{\partial y_k} \times \frac{\partial y_k}{\partial s_k} \times \frac{\partial s_k}{\partial s_t}$$

Let's define $z_t = W_{IH} x_t + W_{HH} s_{t-1}$
Thus,

$$z_k = W_{IH} x_k + W_{HH} s_{k-1}$$
$$s_k = \sigma(z_k)$$
$$y_k = W_{HO} s_k$$

Thus,

$$\frac{\partial y_k}{\partial s_k} = W_{HO} \tag{1}$$

$$\frac{\partial s_k}{\partial z_k} = \sigma(z_k)(1 - \sigma(z_k)) \tag{2}$$

$$\frac{\partial z_k}{\partial W_{IH}} = x_k \tag{3}$$

$$\frac{\partial y_k}{\partial W_{HH}} = y_{k-1} \tag{4}$$

$$\frac{\partial z_k}{\partial s_{k-1}} = W_{HH} \tag{5}$$

$$\frac{\partial s_k}{\partial s_{k-1}} = \frac{\partial s_k}{\partial z_k} \frac{\partial z_k}{\partial s_{k-1}} = \sigma(z_k)(1 - \sigma(z_k)) W_{HH} \tag{6}$$

Let's now consider $\frac{\partial s_k}{\partial s_t}$:
s_k depends on $s_{k-1}, s_{k-2}, \ldots s_1$. And hence:

$$\frac{\partial s_k}{\partial s_t} = 0 \ \forall \ k < t$$

For $k \geq t$:

$$\frac{\partial s_k}{\partial s_t} = \frac{\partial s_k}{\partial s_{k-1}} \times \frac{\partial s_{k-1}}{\partial s_{k-2}} \times \frac{\partial s_{k-2}}{\partial s_{k-3}} \times \ldots \frac{\partial s_{k-(k-t)+1}}{\partial s_{k-(k-t)}}$$

Thus, consider a special case of $t = T$:

$$\nabla_{s_T} L = \sum_{k=T}^{T} \frac{\partial L}{\partial y_k} \times \frac{\partial y_k}{\partial s_k} \times \frac{\partial s_k}{\partial s_t}$$
$$= \frac{\partial L}{\partial y_T} \times \frac{\partial y_T}{\partial s_T}$$
$$= (y_T - \hat{y_T}) W_{HO}$$

Thus,

$$\boxed{\nabla_{S_T} L = (y_T - \hat{y_T}) W_{HO}}$$

Let's consider $\nabla_{s_t} L$ and $\nabla_{s_{t+1}} L$:

$$\nabla_{s_{t+1}} L = \sum_{k=t+1}^{T} \frac{\partial L}{\partial y_k} \times \frac{\partial y_k}{\partial s_k} \times \frac{\partial s_k}{\partial s_t}$$

$$\nabla_{s_t} L = \sum_{k=t}^{T} \frac{\partial L}{\partial y_k} \times \frac{\partial y_k}{\partial s_k} \times \frac{\partial s_k}{\partial s_t}$$

$$\implies \nabla_{s_t} L = \nabla_{s_{t+1}} L + \frac{\partial L}{\partial y_t} \times \frac{\partial y_t}{\partial s_t} \times \frac{\partial s_t}{\partial s_t}$$

$$\implies \nabla_{s_t} L = \nabla_{s_{t+1}} L + (y_t - \hat{y}_t) W_{HO}$$

Thus,

$$\boxed{\nabla_{s_t} L = \nabla_{s_{t+1}} L + (y_y - \hat{y}_t) W_{HO}}$$

Problem 1: (c)

$$\nabla_{W_{IH}} L = \sum_{k=1}^{T} \left(\frac{\partial L}{\partial y_k} \times \frac{\partial y_k}{\partial s_k} \right) \times \frac{\partial s_k}{\partial z_k} \times \frac{\partial z_k}{\partial W_{IH}}$$

$$= \sum_{k=1}^{T} \nabla_{s_k} L \times \sigma(z_k)(1 - \sigma(z_k)) \times x_k$$

Thus,

$$\nabla_{W_{IH}} L = \sum_{k=1}^{T} \nabla_{s_k} L \times \sigma(z_k)(1 - \sigma(z_k)) \times x_k$$
$$\text{where } \nabla_{s_k} L = \nabla_{s_{k+1}} L + (y_k - \hat{y}_k) W_{HO}$$
$$\text{and } z_k = W_{IH} x_k + W_{HH} s_{k-1}$$
$$\text{boundary condition } \nabla_{S_T} L = (y_T - \hat{y}_T) W_{HO}$$

$$\nabla_{W_{HH}} L = \sum_{k=1}^{T} \left(\frac{\partial L}{\partial y_k} \times \frac{\partial y_k}{\partial s_k} \right) \times \frac{\partial s_k}{\partial z_k} \times \frac{\partial z_k}{\partial W_{HH}}$$

$$= \sum_{k=1}^{T} \nabla_{s_k} L \times \sigma(z_k)(1 - \sigma(z_k)) \times s_{k-1}$$

Thus,

$$\nabla_{W_{IH}} L = \sum_{k=1}^{T} \nabla_{s_k} L \times \sigma(z_k)(1 - \sigma(z_k)) \times s_{k-1}$$
$$\text{where } \nabla_{s_k} L = \nabla_{s_{k+1}} L + (y_k - \hat{y}_k) W_{HO}$$
$$\text{and } z_k = W_{IH} x_k + W_{HH} s_{k-1}$$
$$\text{boundary condition } \nabla_{S_T} L = (y_T - \hat{y}_T) W_{HO}$$

$$\nabla_{W_{HO}}L = \sum_{k=1}^{T} \frac{\partial L}{\partial y_k} \times \frac{\partial y_k}{\partial W_{HO}}$$

$$= \sum_{k=1}^{T} (y_k - \hat{y_k})s_k$$

Thus,

$$\nabla_{W_{HO}}L = \sum_{k=1}^{T} (y_k - \hat{y_k})s_k$$

Problem 1: (d)

Leaky hidden units:

$$s_t = (1 - \tau)s_{t-1} + \tau\sigma(z_t)$$

Thus,

$$\frac{\partial s_t}{\partial s_{t-1}} = 1 - \tau + \tau\sigma(z_t)(1 - \sigma(z_t))W_{HH}$$

For $\nabla_{W_{IH}}L$,

$$\nabla_{W_{IH}}L = \sum_{k=1}^{T} \big(\frac{\partial L}{\partial y_k} \times \frac{\partial y_k}{\partial s_k} \big) \times \frac{\partial s_k}{\partial z_k} \times \frac{\partial z_k}{\partial W_{IH}}$$

$$= \sum_{k=1}^{T} \nabla_{s_k}L \times \sigma(z_k)(1 - \sigma(z_k)) \times x_k$$

Problem 2

Problem 2: (a)

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

where

$$\tilde{\mu_k} =$$

Problem 3

Given:
$$p(x_i) = \begin{cases} \pi + (1-\pi)e^{-\lambda} & x_i = 0 \\ (1-\pi)\frac{\lambda^{x_i}e^{-\lambda}}{x_i!} & x_i > 0 \end{cases}$$

Alternatively:

$$X_i = \begin{cases} 0 & \text{probability} = \pi + (1-\pi)e^{-\lambda} \\ x_i & \text{probability} = (1-\pi)\frac{\lambda^{x_i}e^{-\lambda}}{x_i!} \end{cases}$$

We define a *latent* variable Z_i for all cases where $X_i = 0$. It is latent because when we observed $X_i = 0$ we do not know if it came out of the 'Poisson' distribution or it came out the 'degenerate' distribution(which has a probability of 1 at point 0.). we cannot observe the following. So X_i comes out of a mixture of a degenerate distribution as follows:

$$Z_i = \begin{cases} 1 & X_i \text{ is from the degenerate distrbution} \\ 0 & \text{otherwise} \end{cases}$$

$$p(X_i = 0, Z_i = 1) = p(Z_i = 1) \times p(X_i = 0|Z_i = 1) = \pi \times 1$$
$$P(X_i = 0, Z_i = 0) = p(Z_i = 0) \times p(X_i = 0|Z_i = 0) = (1-\pi)e^{-\lambda} \times 1$$

$$L(Complete) = \prod_{x_i=0} \pi^{Z_i} \times ((1-\pi)e^{-\lambda})^{1-Z_i} \times \prod_{x_i>0} (1-\pi)e^{\frac{\lambda_i^x e^{-\lambda}}{x_i!}} \quad (7)$$

$$\log L = \sum_{I(x_i=0)} z_i \log(\pi) + (1-z_i)\big(\log(1-\pi) - \lambda\big) + \sum_{I(x_i>0)} \big(\log(1-\pi) + x_i \log(\lambda_i) - \lambda - \log(x_i!)\big) \quad (8)$$

E step:

$$Q(\theta, \theta_0) = \sum_{I(x_i=0)} E_{P(Z|X)}[z_i] \log(\pi) + (1 - E_{P(Z|X)}[z_i])\big(\log(1-\pi) - \lambda\big)$$
$$+ \sum_{I(x_i>0)} \big(\log(1-\pi) + x_i \log(\lambda_i) - \lambda - \log(x_i!)\big)$$

$$E_{P(Z|X_i)}[z_i] = 0 \times p(Z_i = 0|X) + 1 \times p(Z_i = 1|X_i = 0)$$
$$= \frac{p(X_i = 0|Z_i = 1)p(Z_i = 1)}{p(X_i = 0|Z_i = 0)p(Z_i = 0) + p(X_i = 0|Z_i = 1)p(Z_i = 1)}$$
$$= \frac{\pi_0}{\pi_0 + (1-\pi_0)e^{-\lambda_0}}$$

Hence,

$$Q(\theta, \theta_0) = \sum_{I(x_i=0)} \frac{\pi_0}{\pi_0 + (1-\pi_0)e^{-\lambda_0}} \log(\pi) + \big(\frac{(1-\pi_0)e^{-\lambda_0}}{\pi_0 + (1-\pi_0)e^{-\lambda_0}}\big)\big(\log(1-\pi) - \lambda\big)$$
$$+ \sum_{I(x_i>0)} \big(\log(1-\pi) + x_i \log(\lambda) - \lambda - \log(x_i!)\big)$$

M step:

$$\frac{\partial Q}{\partial \lambda} = 0$$

$$= \sum_{I(x_i=0)} (1 - E[z_i])(-1) + \sum_{I(x_i>0)} \left(\frac{x_i}{\lambda} - 1\right) = 0$$

$$\implies \hat{\lambda} = \frac{\sum_{I(x_i>0)} x_i}{n - \sum_{I(x_i=0)} E[z_i]}$$

$$\hat{\lambda} = \frac{\sum_{I(x_i>0)} x_i}{n - \sum_{I(x_i=0)} \hat{z}_i}$$

$$\text{where } \hat{z} = \frac{\pi_0}{\pi_0 + (1 - \pi_0)e^{-\lambda_0}}$$

$$\frac{\partial Q}{\partial \pi} = 0$$

$$= \sum_{I(x_i=0)} \left(\frac{E[z_i]}{\pi} - \frac{1 - E[z_i]}{1 - \pi}\right) - \sum_{I(x_i>0)} \frac{1}{1 - \pi} = 0$$

$$= \sum_{I(x_i=0)} \left(\frac{E[z_i]}{\pi} + \frac{E[z_i]}{1 - \pi}\right) - \frac{n}{1 - \pi} = 0$$

$$\implies \hat{\pi} = \sum_{I(x_i=0)} \frac{\hat{z}_i}{n}$$

Problem 4

Problem 4.2

As seen from Figure 2, k-means algorithm fails to separate the two circles. This happens because of the failure of the following assumption in this case: the dataset is linearly separable. The circular dataset is not really separable, and hence the clusters returned by k-means, have a linear boundary(making two halves of the circle)
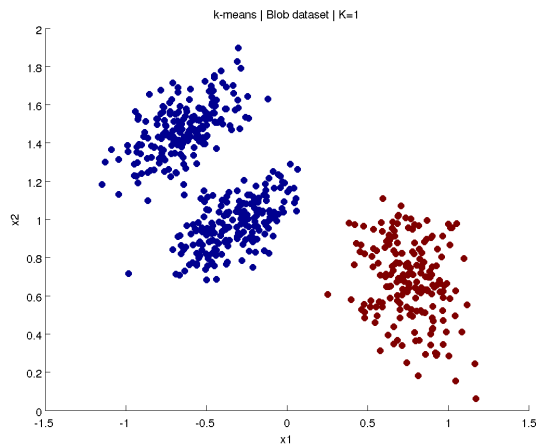
Figure 1: Blob Dataset $k = 2$
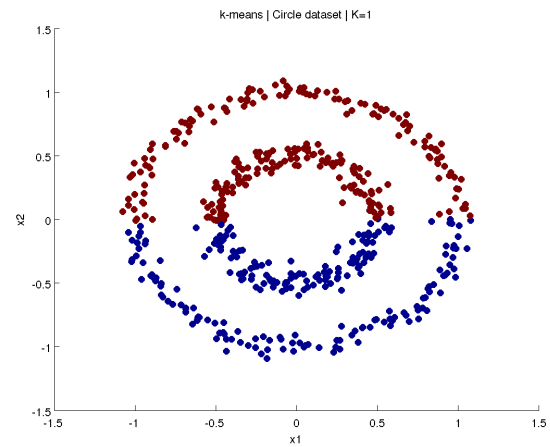
Figure 2: Circle Dataset $k = 2$
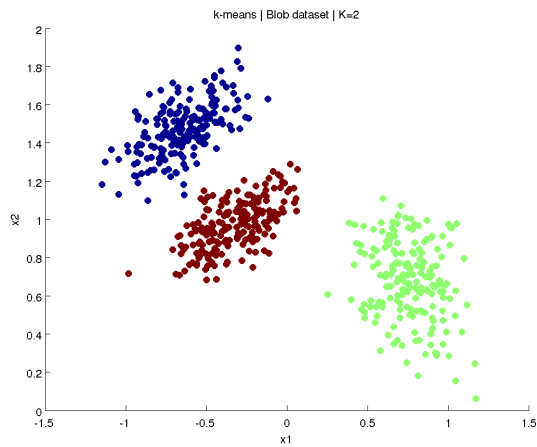
Figure 3: Blob Dataset $k = 3$
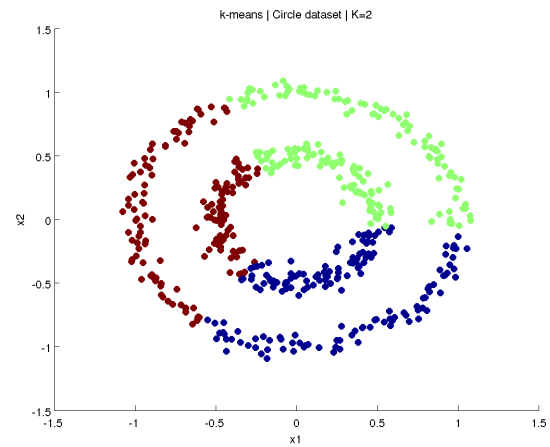
Figure 4: Circle Dataset $k = 3$
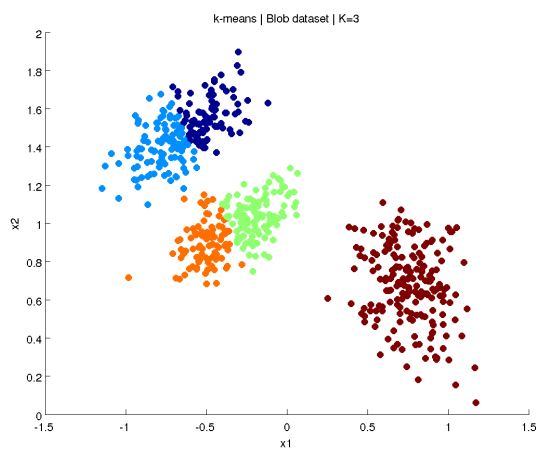
Figure 5: Blob Dataset $k = 5$
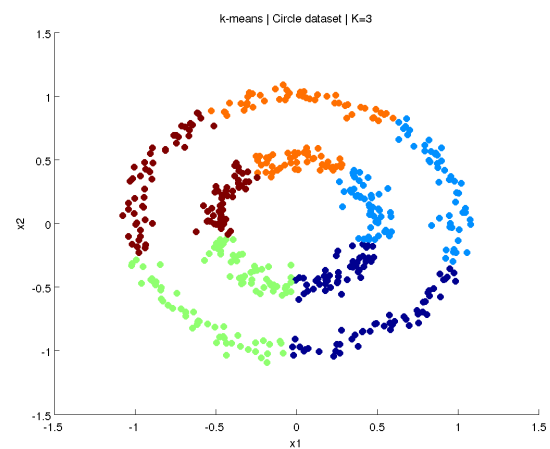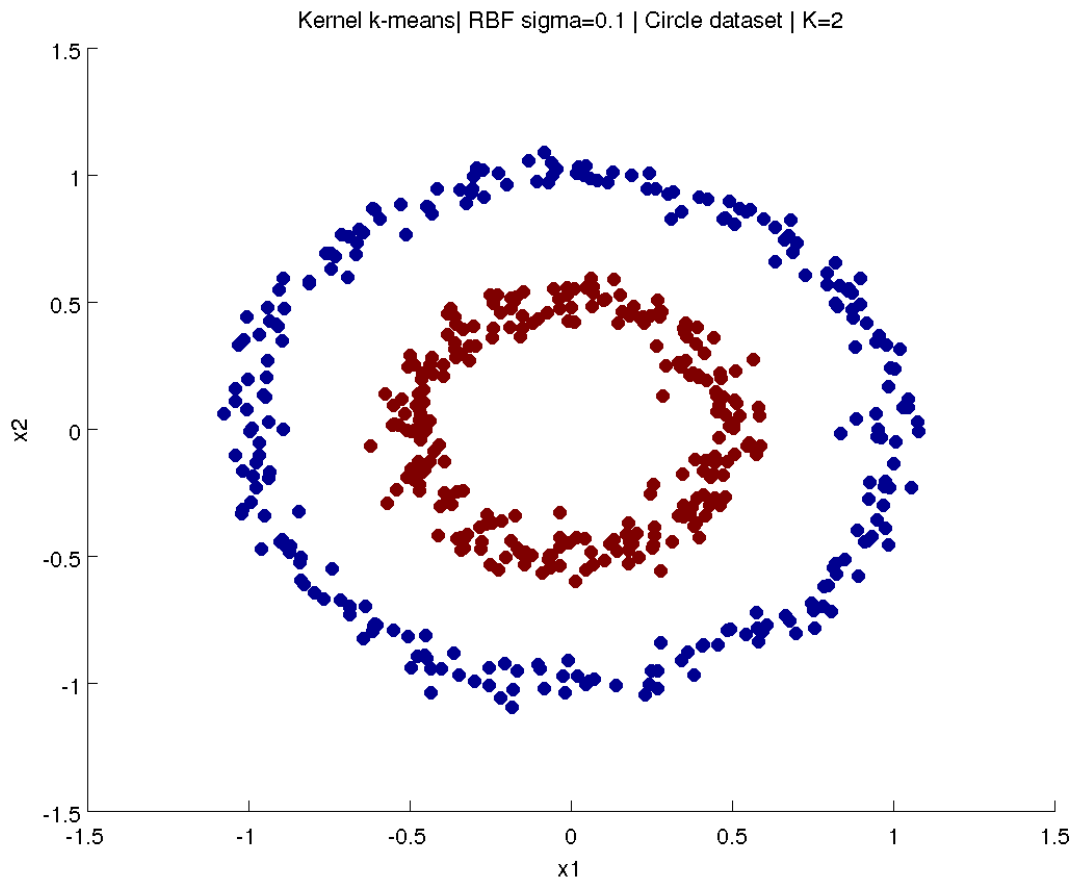

Figure 6: Circle Dataset $k = 5$

Figure 7: Problem 4.3 Kernel k-means with RBF kernel creates separate clusters

Problem 4.3

Choice of Kernel = RBF:
$$K(x, x') = \exp(-\frac{||x - x'||^2}{2\sigma^2})$$
with $\sigma = 0.1$

Problem 4.4(a)

Problem 4.4(b)