EEE – 485

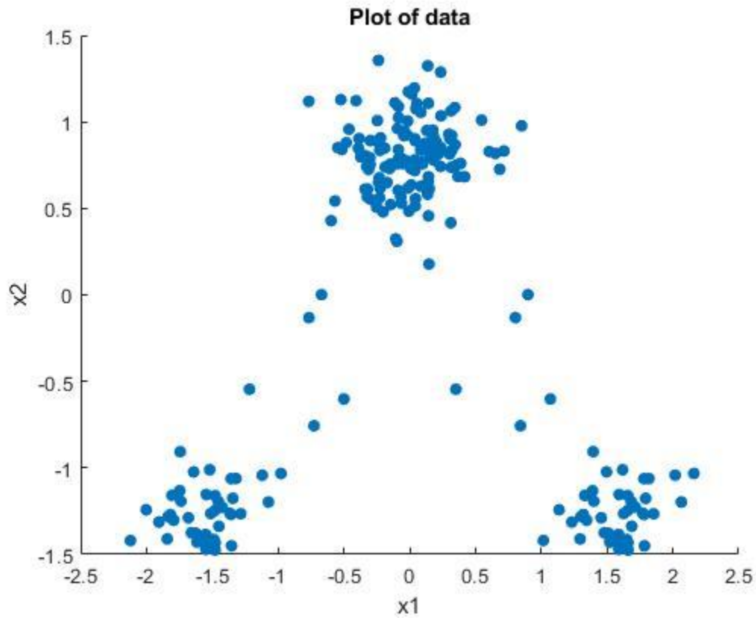
Statistical Learning and Data Analytics

Problem Set # 4

Kerem Ayöz

21501569 - CS

a-) Plot of standardized data



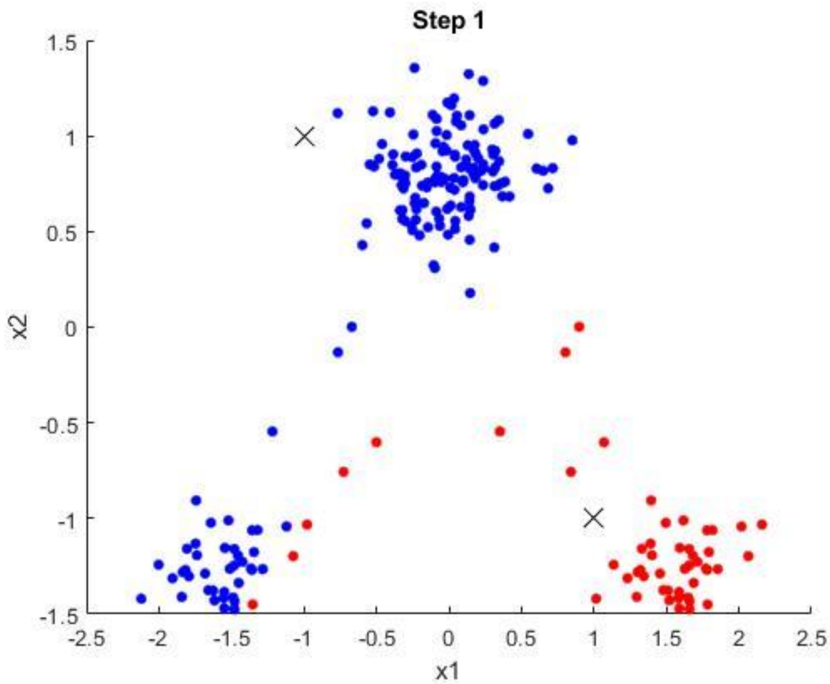
b-) Code for K-Means Clustering

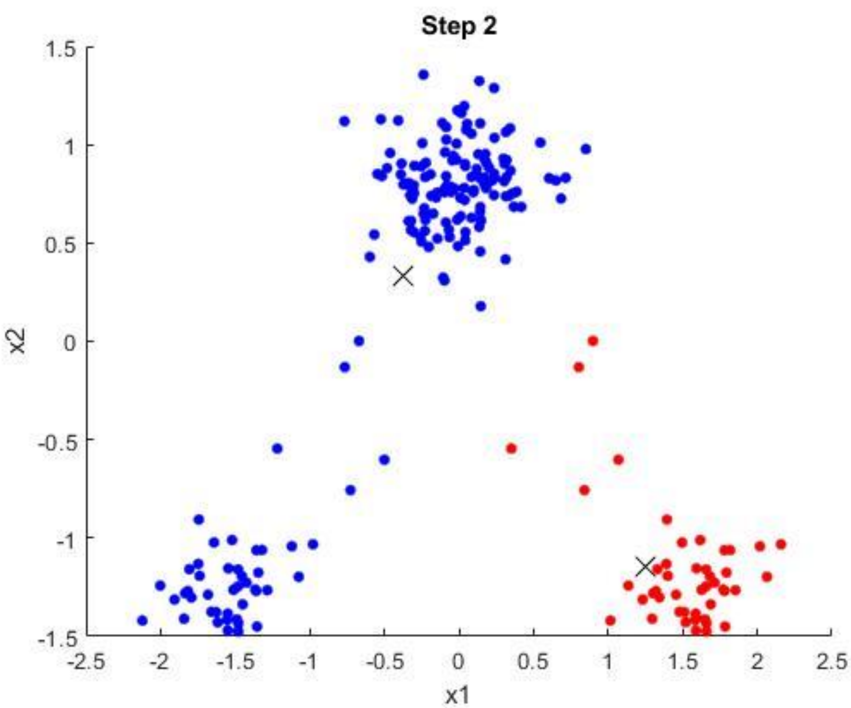
```
load("ps4q1.mat");
% Centralize
for i = (1:2)
    data(:,i) = data(:,i) - mean(data(:,i));
end
% Standardize
for i = (1:2)
    data(:,i) = data(:,i) / std(data(:,i));
end

figure;
scatter(data(:,1),data(:,2), 'filled');
title("Plot of data");
xlabel("x1");
ylabel("x2");

% Parameters
K = 2;
means = [-1,1;1,-1];
norm = "l1";

% Clustering
[means, classes] = k_means_clustering(K, data, means, norm);

% Plot the results
plot_clusters(data, classes, K, means, "Result");
```

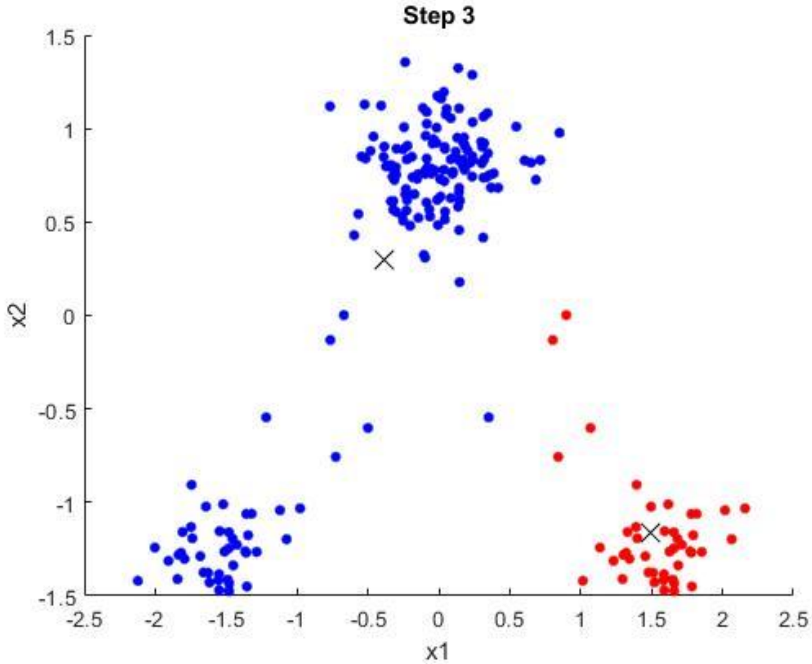
```
% K-Means Clustering
function [result_means, result_classes] = k_means_clustering(K, data, means,
norm)
    for e = (1:100)
        temp = means;

        % Expectation
        classes = zeros(length(data),1);
        for i = (1:length(data))
            loss_i = zeros(1,K);
            for j = (1:K)
                points = [data(i,:);means(j,:)];
                loss_i(j) = loss(points, norm);
            end
            classes(i) = find(loss_i == min(loss_i(:)));
        end

        figure;
        plot_clusters(data, classes, K, means, char(cellstr(num2str(e, 'Step
%d'))));

        % Maximization
        for i = (1:K)
            means(i,:) = mean(data((classes(:) == i),:));
        end
        if means == temp
            break
        end
    end
    result_classes = classes;
    result_means = means;
end

% Distance Function
function distance = loss(points, form)
    if form == "l1"
        distance = pdist(points,'euclidean');
    elseif form == "l2"
        distance = pdist(points,'minkowski', 1);
    end
end

% Plots the data points with clusters
function plot_clusters(data, classes, K, means, p_title)
    colors = ['b', 'r', 'g', 'y', 'k', 'c', 'm'];
    for i = (1:K)
        title(p_title);
        xlabel("x1");
        ylabel("x2");
        scatter(data((classes == i),1), data((classes == i),2), length(data),
colors(mod(i,7)), '.')
        hold on;
        scatter(means(i,1), means(i,2), 150, 'k', 'x')
        hold on;
    end
end
```

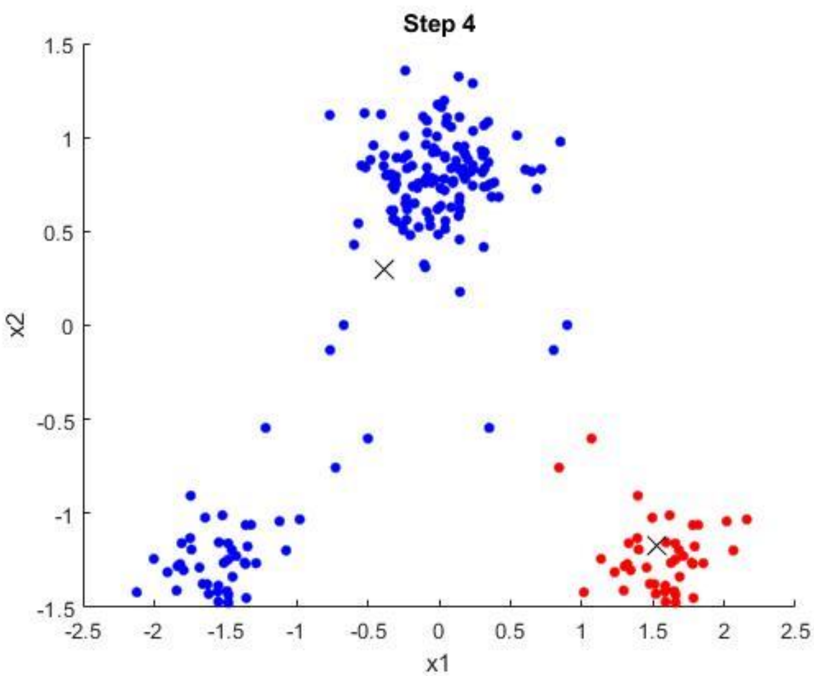
c-)

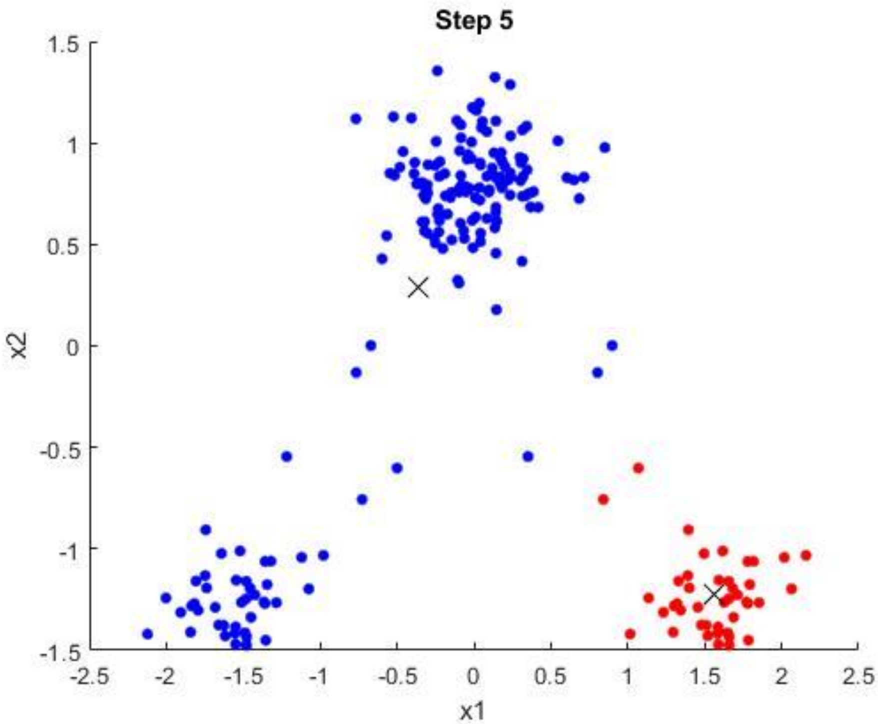
- For K=2 and l1 (Euclidian Distance) norm, each step is plotted.



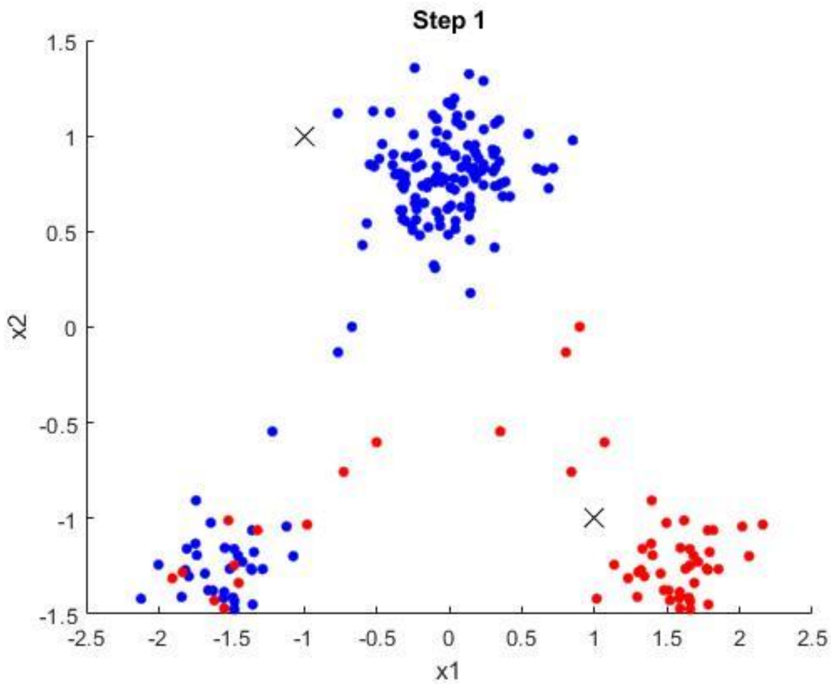


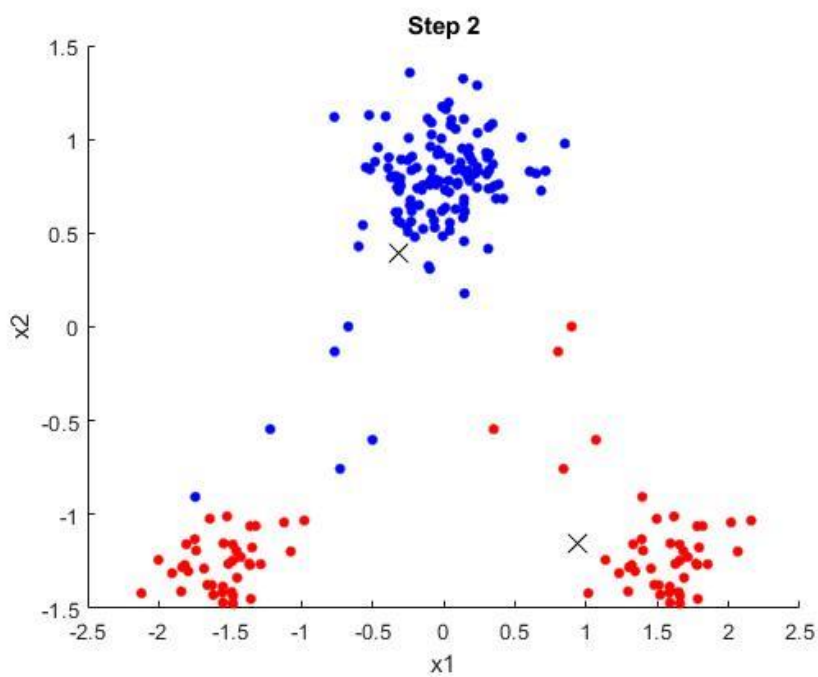

Step 3
1.5
1
0.5
0
-0.5
-1
-1.5
x2
-2.5 -2 -1.5 -1 -0.5 0 0.5 1 1.5 2 2.5
x1

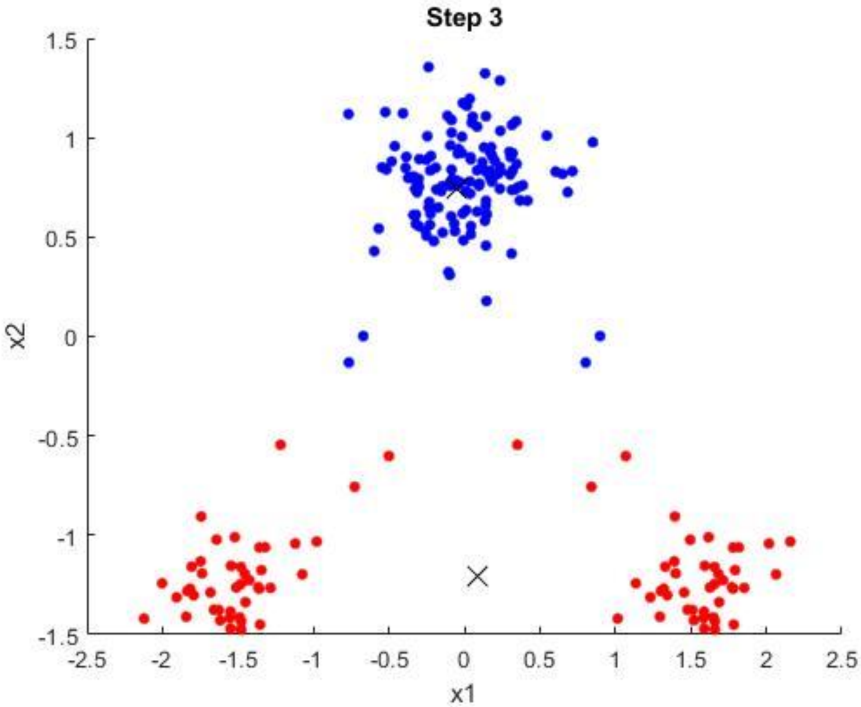

Step 4
1.5
1
0.5
0
-0.5
-1
-1.5
x2
-2.5 -2 -1.5 -1 -0.5 0 0.5 1 1.5 2 2.5
x1

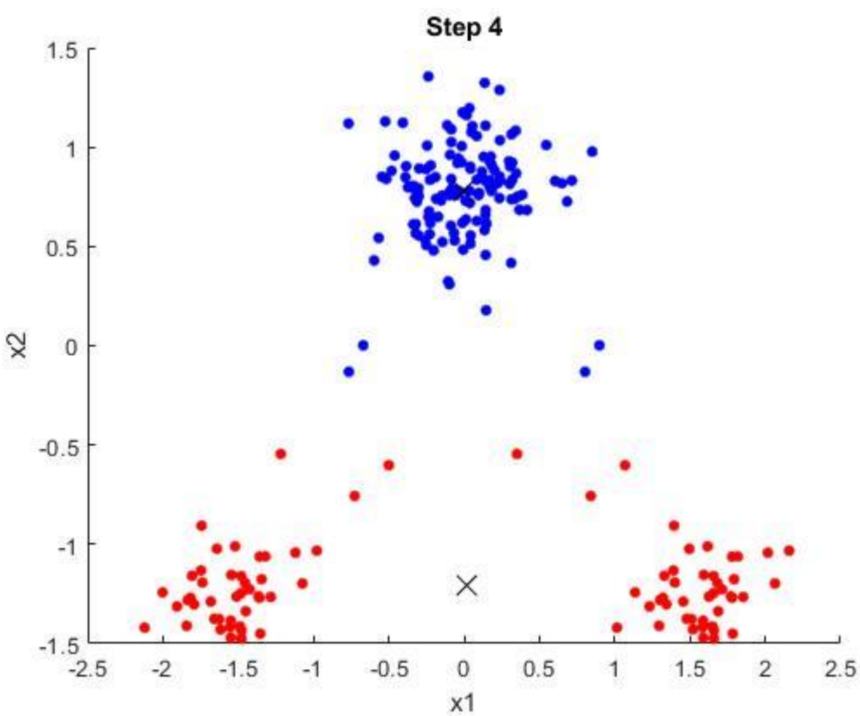

Step 5
x2
1.5
1
0.5
0
-0.5
-1
-1.5
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
x1

- For K=2 and l2 (Absolute Distance) norm, each step is plotted.



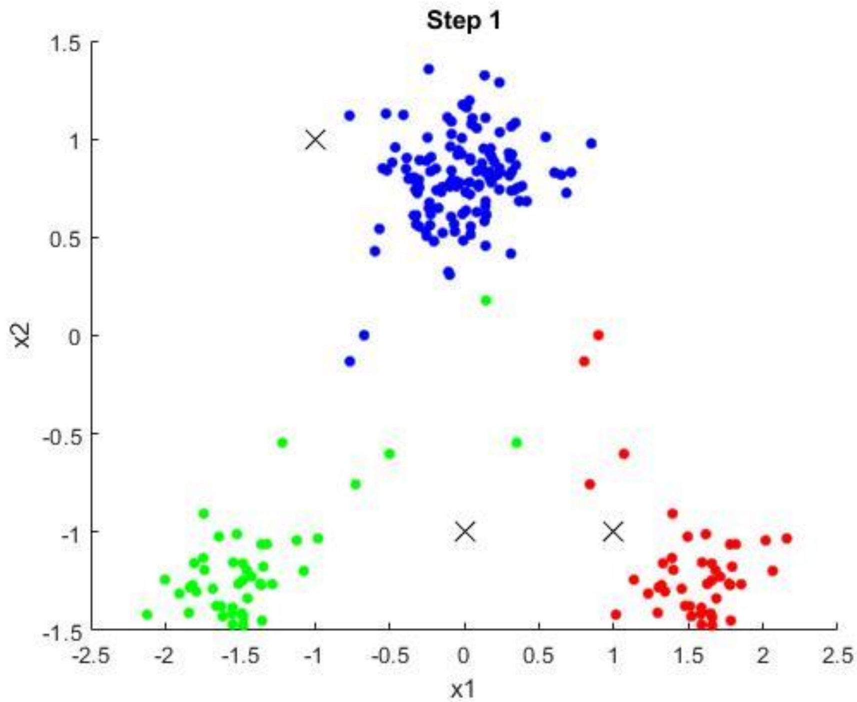


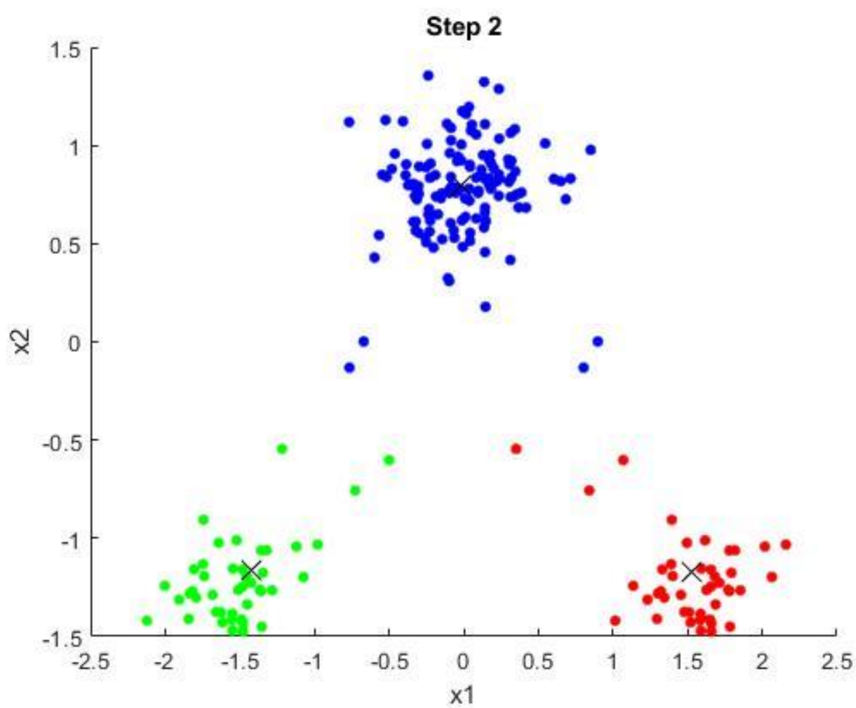

Step 3
x2
x1
1.5
1
0.5
0
-0.5
-1
-1.5
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5

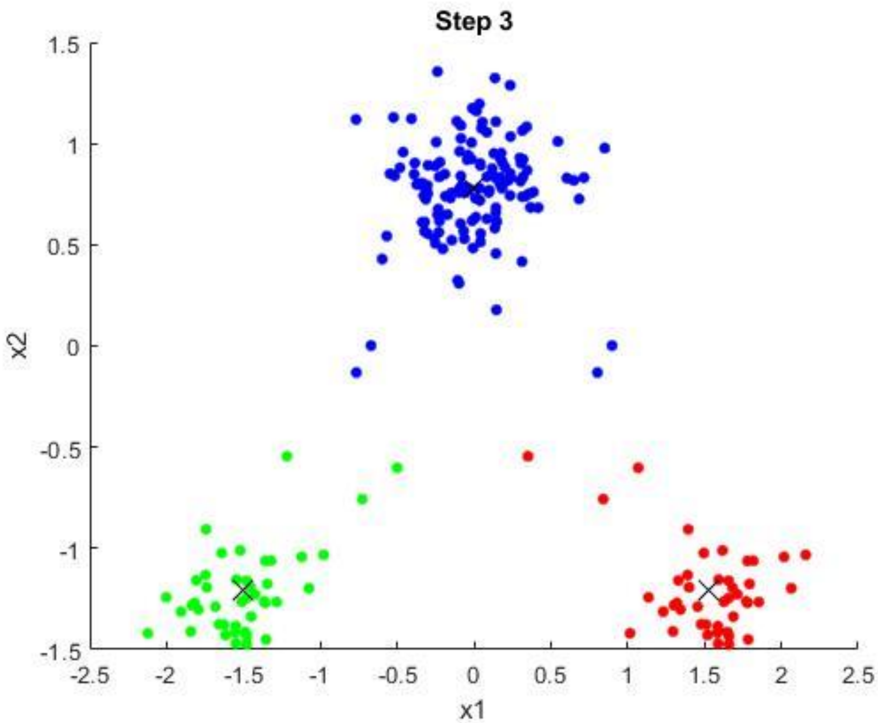

Step 4
x2
x1
1.5
1
0.5
0
-0.5
-1
-1.5
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5

d-)

- For K=3 and l1 (Euclidian Distance) norm, each step is plotted.

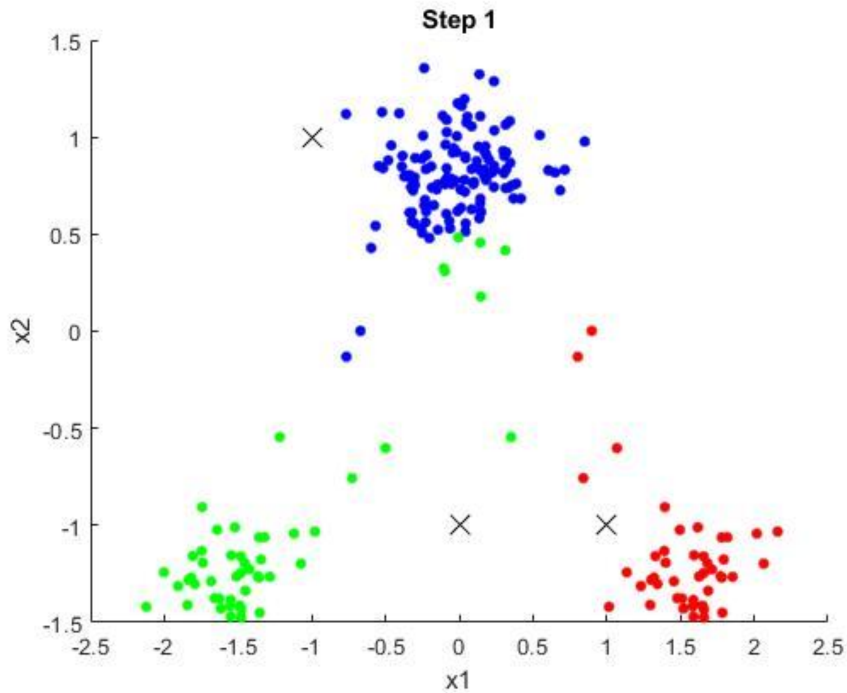


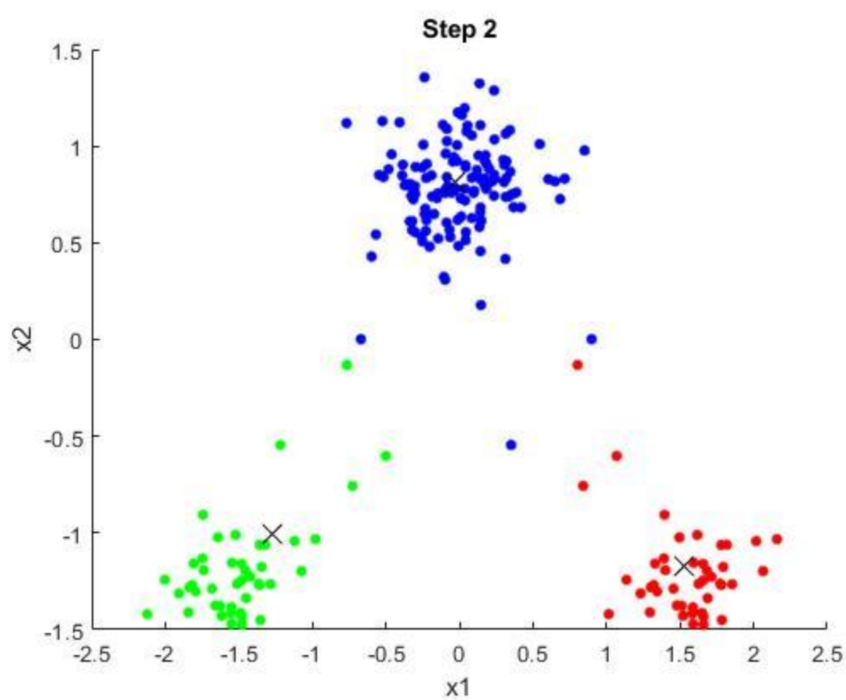


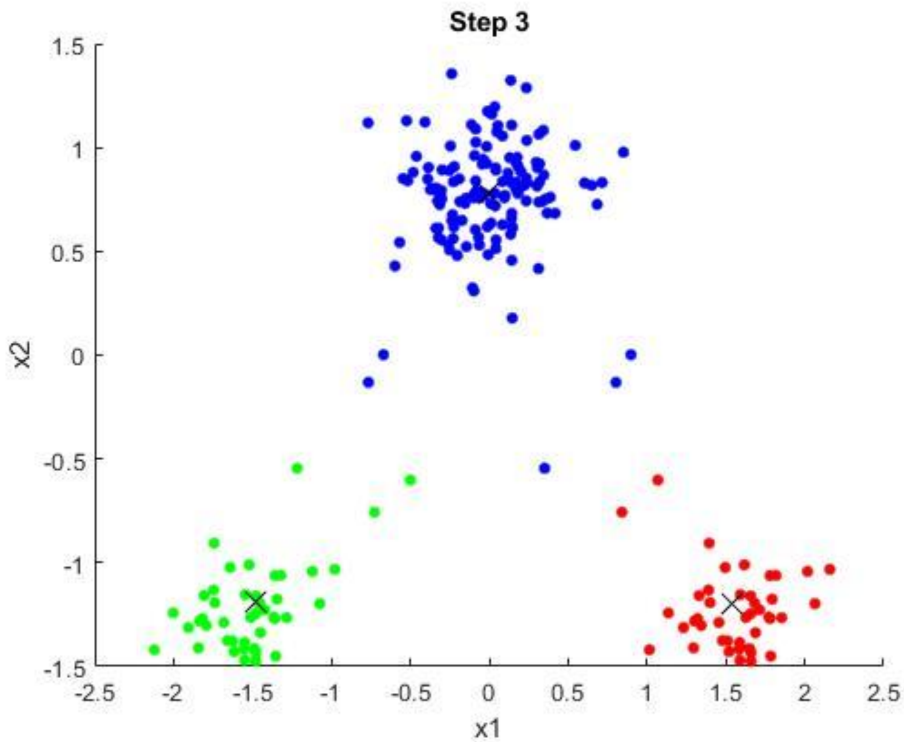


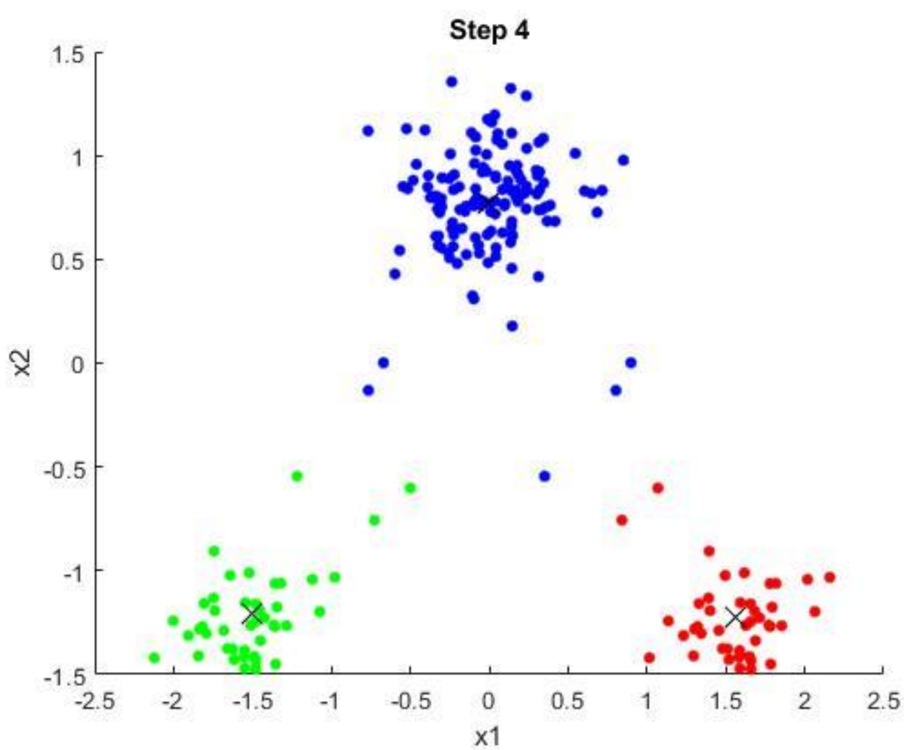
Step 3
1.5
1
0.5
0
-0.5
-1
-1.5
x2
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
x1

- For K=3 and l2 (Absolute Distance) norm, each step is plotted.





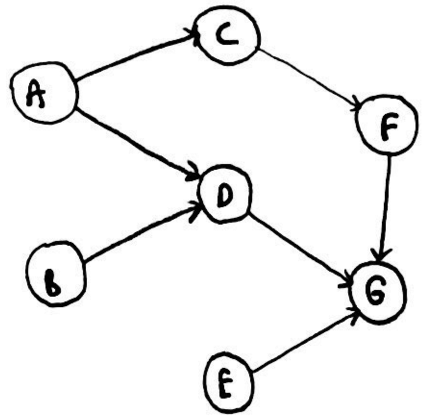

Step 3
1.5
1
0.5
0
-0.5
-1
-1.5
x2
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
x1


Step 4
1.5
1
0.5
0
-0.5
-1
-1.5
x2
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
x1

e-) As I observed from the plots of K-Means Clustering algorithms steps, l2 loss is better for that dataset, since for K=2 it divides the dataset more uniformly than the l1 norm. Also for K=3 even there are little differences in the results, l2 norm gives better results.

f-) K=3 is more suitable since when we take K=2 we get unintuitive results, for l1 norm algorithm clusters too many elements into class blue, in l2 norm algorithm clusters too many elements into class red so it cannot divide the data uniformly. However, when K=3, we get 3 nearly equal sized clusters and that is what we aim when applying the K-Means Clustering algorithm.

2



a.) $p(a,b,c,d,e,f,g) = p(a)\,p(b)\,p(e)\,p(c|a)\,p(f|c)\,p(d|a,b)\,p(g|d,e,f)$

b.) $p(a,e) \overset{?}{=} p(a)\,p(e)$

$$p(a,e) = \sum_{b,c,d,f,g} p(a,b,c,d,e,f,g) = p(a)p(e) \underbrace{\sum_{b,c,d,f,g} p(c|a)\,p(f|c)\,p(b)\,p(d|a,b)\,p(g|d,e,f)}_{\text{Not equal to 1}}$$

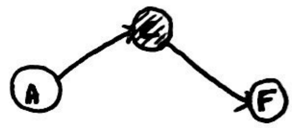
Thus a and e are not independent $\boxed{A \not\perp\!\!\!\perp E}$

c.) We have 2 paths

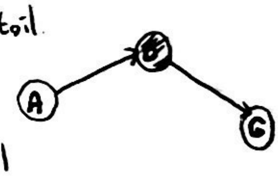
$P_1: A-C-F-G-E$

$P_2: A-D-G-E$

- P_1 is blocked since C is observed and head-tail.

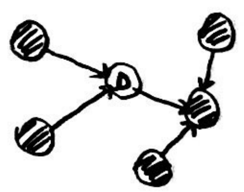


- P_2 is blocked since D is observed and head-tail



Thus A and E are conditionally independent given C, D $\boxed{A \perp\!\!\!\perp E \mid C,D}$

d.) A and B are parents of D so they are in its Markov Blanket.
G is child of D so it is in its Markov Blanket.
E and F are co-parent of D so they are in its Markov Blanket.



e.)

③ a) • $p(h|x) = \dfrac{p(h,x)}{p(x)} = \dfrac{e^{+\sum\sum w_{ij}h_i x_j + \sum b_i x_i + \sum c_i h_i}}{Z\; p(x)} = \dfrac{e^{h^T W x + b^T x + c^T h}}{p(x)\; Z}$

$$= \frac{e^{h^T W x}\, \cancel{e^{b^T x}}\, e^{c^T h}}{\cancel{e^{b^T x}}\, e^{\sum_{i=1}^{n} \log(1+\exp(c_i + \sum_{j=1}^{m} w_{ij} x_j))}\, \frac{1}{\cancel{Z}}\, \cancel{Z}}$$

$$p(h|x) = \frac{e^{h^T W x}\, e^{c^T h}}{\underbrace{e^{\sum_{i=1}^{n} \log(1+\exp(c_i + \sum_{j=1}^{m} w_{ij} x_j))}}_{\text{call it } \varphi}}$$

independent from h

$$p(h_i = 1|x) = \frac{\tilde{p}(h_i = 1|x)}{\tilde{p}(h_i = 0|x) + \tilde{p}(h_i = 1|x)} = \frac{\frac{e^{Wx}\, e^{c^T}}{\cancel{\varphi}}}{\frac{e^{0}}{\cancel{\varphi}} + \frac{e^{Wx}\, e^{c^T}}{\cancel{\varphi}}} = \frac{e^{Wx + c^T}}{1 + e^{Wx + c^T}}$$

Then $p(h_i = 1|x) = \dfrac{e^{Wx + c^T}}{1 + e^{Wx + c^T}} = \dfrac{1}{1 + e^{-Wx - c^T}} = \phi(Wx + c^T) = \phi\left(\sum_{j=1}^{m} w_{ij} x_j + c_i\right)$

• $p(x|h) = \dfrac{p(h,x)}{p(h)} = \dfrac{e^{h^T W x + b^T x + c^T h}}{p(h)\; Z} = \dfrac{e^{h^T W x}\, e^{b^T x}\, \cancel{e^{c^T h}}}{\cancel{e^{c^T h}}\, \underbrace{e^{\sum_{j=1}^{m} \log(1+\exp(b_j + \sum_{i=1}^{n} w_{ij} h_i))}}_{\text{call it } \tilde{\varphi}}\, \frac{1}{\cancel{Z}}\, \cancel{Z}}$

independent from x

$$= \frac{e^{h^T W x}\, e^{b^T x}}{\tilde{\varphi}}$$

$$p(x_i = 1|h) = \frac{p(x_i = 1, h)}{p(x_i = 0, h) + p(x_i = 1, h)} = \frac{\frac{e^{h^T W + b^T}}{\cancel{\tilde{\varphi}}}}{\frac{e^{0}}{\cancel{\tilde{\varphi}}} + \frac{e^{h^T W + b^T}}{\cancel{\tilde{\varphi}}}} = \frac{e^{h^T W + b^T}}{1 + e^{h^T W + b^T}}$$

Then $p(x_i = 1|h) = \dfrac{e^{h^T W + b^T}}{1 + e^{h^T W + b^T}} = \dfrac{1}{1 + e^{-h^T W - b^T}} = \phi(h^T W + b^T) = \phi\left(\sum_{i=1}^{n} w_{ij} h_i + b_j\right)$

b.) $$p(x) = \sum_h p(x,h) = \sum_h \frac{\exp(-E(x,h))}{\sum_{x',h'} \exp(-E(x',h'))} = \frac{1}{Z} \sum_h \exp(-E(h,x))$$

$$p(x) = \frac{1}{Z} \sum_h e^{h^T W x + b^T x + c^T h}$$

$$= \frac{1}{Z} \sum_h e^{\sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} h_i x_j + \underbrace{\sum_{j=1}^{m} b_j x_j}_{b^T x} + \sum_{i=1}^{n} c_i h_i}$$

$$= \frac{1}{Z} e^{b^T x} \cdot \sum_{h \in \{0,1\}^n} e^{\sum_{i=1}^{n} h_i \left(c_i + \sum_{j=1}^{m} w_{ij} x_j\right)}$$

$$= \frac{1}{Z} e^{b^T x} \cdot \sum_{h \in \{0,1\}^n} \prod_{i=1}^{n} e^{h_i \left(c_i + \sum_{j=1}^{m} w_{ij} x_j\right)}$$

$$= \frac{1}{Z} e^{b^T x} \prod_{i=1}^{n} \sum_{h \in \{0,1\}^n} e^{h_i \left(c_i + \sum_{j=1}^{m} w_{ij} x_j\right)}$$

$$= \frac{1}{Z} e^{b^T x + \sum_{i=1}^{n} \log\left(1 + e^{\left(c_i + \sum_{j=1}^{m} w_{ij} x_j\right)}\right)}$$

c) $I(\theta) = \log p(x^T|\theta)$ where $\theta = \{w_{ij}, b_j, c_i\}$

$$I(\theta) = \log \sum_h e^{-E(x^t,h)} - \log \sum_{x,h} e^{-E(x,h)}$$

$$= \log \sum_h e^{\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} h_i x_j^t + \sum_{j=1}^{m} b_j x_j^t + \sum_{i=1}^{n} c_i h_i} - \log \sum_{x,h} e^{\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} h_i x_j + \sum_{j=1}^{m} b_j x_j + \sum_{i=1}^{n} c_i h_i}$$

Let $\tilde{\theta} \in \{w_{ij}, b_j, c_i\}$

$$\frac{\partial I(\theta)}{\partial \tilde{\theta}} = \frac{\partial}{\partial \tilde{\theta}}\left[\log \sum_h e^{-E(x^t,h)}\right] - \frac{\partial}{\partial \tilde{\theta}}\left[\log \sum_{x,h} e^{-E(x,h)}\right]$$

$$= \frac{1}{\sum_h e^{-E(x^t,h)}} \cdot \sum_h e^{-E(x^t,h)} \cdot \frac{\partial E(x^t,h)}{\partial \tilde{\theta}} + \frac{1}{\sum_{x,h} e^{-E(x,h)}} \sum_{x,h} e^{-E(x,h)} \frac{\partial E(x,h)}{\partial \tilde{\theta}}$$

$$= -\sum_h p(h|x) \frac{\partial E(x^t,h)}{\partial \tilde{\theta}} + \sum_{x,h} p(x,h) \frac{\partial E(x,h)}{\partial \tilde{\theta}}$$

• $$\frac{\partial I(\theta)}{\partial w_{ij}} = \underbrace{+\sum_h p(h|x) \cdot (+h_i x_j^t)}_{A} + \underbrace{\sum_{x,h} p(x,h)(-h_i x_j)}_{B}$$

$$A = \sum_h p(h|x) h_i x_j^t = \sum_{h_i} \sum_{h_{-i}} p(h_i|x^t) p(h_{-i}|x^t) h_i x_j^t = \underbrace{\sum_{h_{-i}} p(h_{-i}|x^t)}_{1} \underbrace{\sum_{h_i} p(h_i|x^t) h_i x_j^t}_{p(h_i=1|x^t)}$$

$$A = 1 \cdot p(h_i=1|x^t) x_j^t$$

$$B = -\sum_{x,h} p(x,h)(h_i x_j) = -\sum_{x,h} p(x) p(h|x)(h_i x_j) = \sum_x p(x) \underbrace{\sum_h p(h|x)(h_i x_j)}_{p(h_i=1|x)}$$

$$B = -\sum_x p(x) p(h_i=1|x) x_j$$

Then $$\frac{\partial I(\theta)}{\partial w_{ij}} = A + B = p(h_i=1|x^t) x_j^t - \sum_x p(x) p(h_i=1|x) x_j$$

$$\bullet\!\!> \quad \frac{\partial I(\theta)}{\partial b_j} = \underbrace{+\sum_h p(h|x^t) \,.\, +x_j^t}_{A} + \underbrace{\sum_{x,h} p(x,h) \,.\, -x_j^t}_{B}$$

$$A = +\sum_h p(h|x^t)\, x_j^t = \sum_{h_i}\sum_{h_{-i}} p(h_i|x^t)\, p(h_{-i}|x^t)\, x_j^t = \underbrace{\sum_{h_{-i}} p(h_{-i}|x^t)}_{1} \underbrace{\sum_{h_i} p(h_i|x^t)}_{1} \,.\, x_j^t$$

$$A = x_j^t$$

$$B = -\sum_{x,h} p(x,h)\, x_j^t = -\sum_{x,h} p(x)\, p(h|x)\, x_j^t = \sum_x p(x) \underbrace{\sum_h p(h|x)}_{1} x_j$$

$$B = -\sum_x p(x)\, x_j$$

$$\text{Then} \quad \frac{\partial I(\theta)}{\partial b_j} = A + B = x_j^t - \sum_x p(x)\, x_j$$

$$\bullet\!\!> \quad \frac{\partial I(\theta)}{\partial c_i} = \underbrace{+\sum_h p(h|x^t)\, h_i}_{A} + \underbrace{\sum_{x,h} p(x,h) \,.\, -h_i}_{B}$$

$$A = \sum_h p(h|x^t)\, h_i = \sum_{h_i}\sum_{h_{-i}} p(h_i|x^t)\, p(h_{-i}|x^t)\, h_i = \underbrace{\sum_{h_{-i}} p(h_{-i}|x^t)}_{1} \underbrace{\sum_{h_i} p(h_i|x^t)\, h_i}_{p(h_i=1|x^t)}$$

$$A = p(h_i=1|x^t)$$

$$B = -\sum_{x,h} p(x,h)\, h_i = -\sum_{x,h} p(x)\, p(h|x)\, h_i = -\sum_x p(x) \underbrace{\sum_h p(h|x)\, h_i}_{p(h_i=1|x)}$$

$$B = -\sum p(x)\, p(h_i=1|x)$$

$$\text{Then} \quad \frac{\partial I(\theta)}{\partial c_i} = A + B = p(h_i=1|x^t) - \sum_x p(x)\, p(h_i=1|x)$$