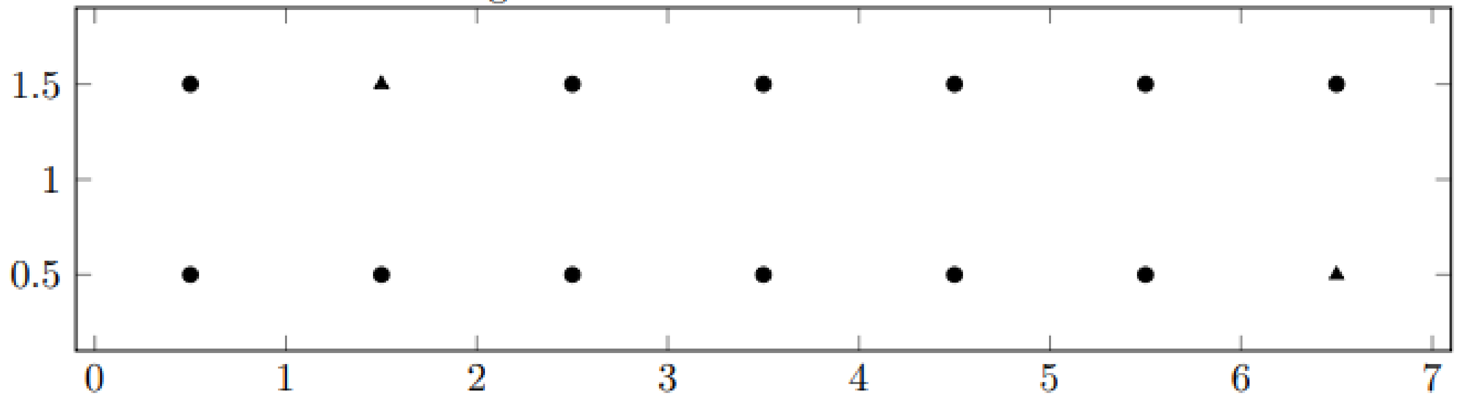
CSCI 567 Discussion Session

1 Clustering

Consider the following dataset. All points are unlabeled and part of the same set. The triangles are used to distinguish two points later.

Suppose that we run the K-means algorithm on this dataset with $K = 2$ and the two points indicated by triangles as the initial centers. When the algorithm converges, there will be two clearly separated clusters. Directly on Figure 1, draw a straight line that separates these two clusters, as well as the centers of these two clusters.

Figure 1: An unlabeled dataset

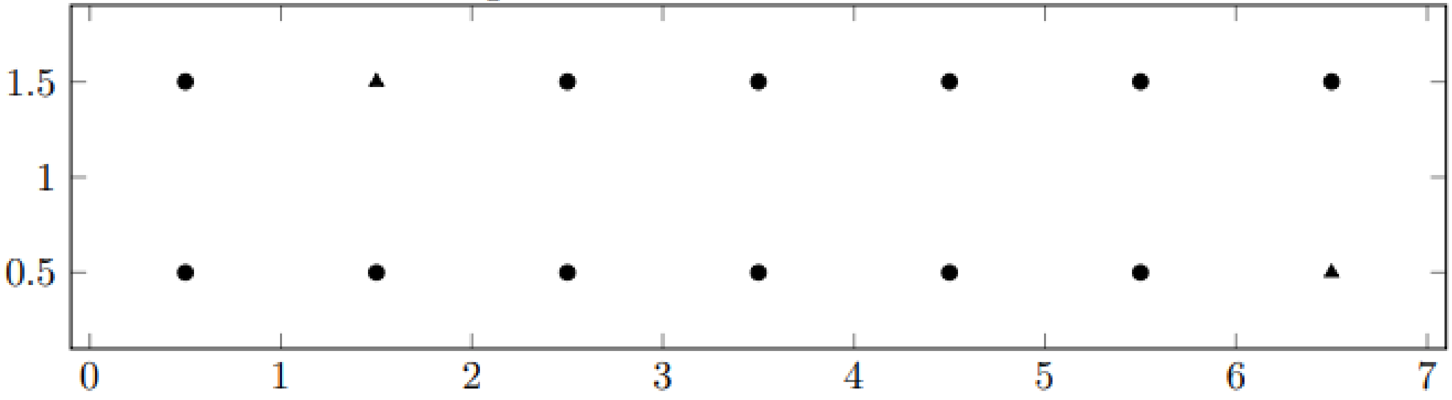


1 Clustering

Consider the following dataset. All points are unlabeled and part of the same set. The triangles are used to distinguish two points later.

Suppose that we run the K-means algorithm on this dataset with $K = 2$ and the two points indicated by triangles as the initial centers. When the algorithm converges, there will be two clearly separated clusters. Directly on Figure 1, draw a straight line that separates these two clusters, as well as the centers of these two clusters.

Figure 1: An unlabeled dataset



- Two initial centers: c1 (1.5, 1.5) and c2 (6.5,0.5).
- K-means algorithm:
  1. Assign each point to the closest cluster.
  2. Updates the center based on the assignment.
  3. Return to step 1 if not converged.

is simply to **assign each x_n to the closest μ_k**, i.e.

$$\gamma_{nk} = \begin{cases} 1, & \text{if } k = \operatorname{argmin}_c \|\boldsymbol{x}_n - \boldsymbol{\mu}_c\|_2^2 \\ 0, & \text{else} \end{cases}$$

is simply **to average the points of each cluster** (hence the name)

$$\boldsymbol{\mu}_k = \frac{\sum_{n:\gamma_{nk}=1} \boldsymbol{x}_n}{|\{n : \gamma_{nk} = 1\}|} = \frac{\sum_n \gamma_{nk} \boldsymbol{x}_n}{\sum_n \gamma_{nk}}$$

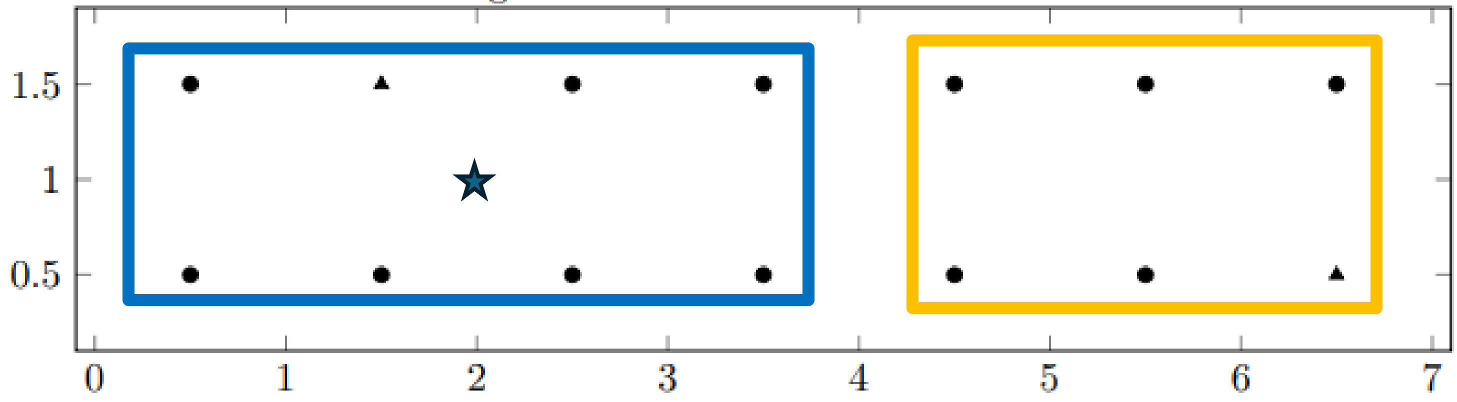
K-means: 10/24 lecture slides page 13-17.

1 Clustering

Consider the following dataset. All points are unlabeled and part of the same set. The triangles are used to distinguish two points later.

Suppose that we run the K-means algorithm on this dataset with $K = 2$ and the two points indicated by triangles as the initial centers. When the algorithm converges, there will be two clearly separated clusters. Directly on Figure 1, draw a straight line that separates these two clusters, as well as the centers of these two clusters.

Figure 1: An unlabeled dataset



- Two initial centers: c1 (1.5, 1.5) and c2 (6.5,0.5).
- Add points to cluster 1.
- Average x-coordinate:

$$\frac{0.5+0.5+1.5+1.5+2.5+2.5+3.5+3.5}{8} = \frac{16}{8} = 2.$$

- Average y-coordinate:

$$\frac{0.5+1.5+0.5+1.5+0.5+1.5+0.5+1.5}{8} = \frac{8}{8} = 1.$$

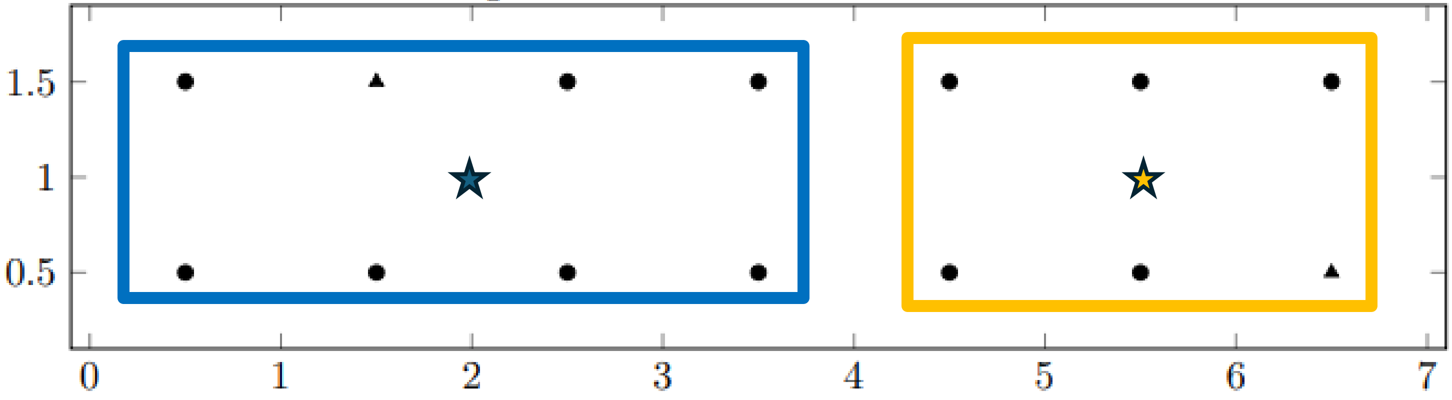
- New center: **(2,1)**

1 Clustering

Consider the following dataset. All points are unlabeled and part of the same set. The triangles are used to distinguish two points later.

Suppose that we run the K-means algorithm on this dataset with $K = 2$ and the two points indicated by triangles as the initial centers. When the algorithm converges, there will be two clearly separated clusters. Directly on Figure 1, draw a straight line that separates these two clusters, as well as the centers of these two clusters.

Figure 1: An unlabeled dataset



- Two initial centers: c1 (1.5, 1.5) and c2 (6.5,0.5).
- Add points to cluster 2.
- Average x-coordinate:

$$\frac{4.5 + 4.5 + 5.5 + 5.5 + 6.5 + 6.5}{6} = \frac{33}{6} = 5.5.$$

- Average y-coordinate:

$$\frac{0.5 + 1.5 + 0.5 + 1.5 + 0.5 + 1.5}{6} = \frac{6}{6} = 1.$$

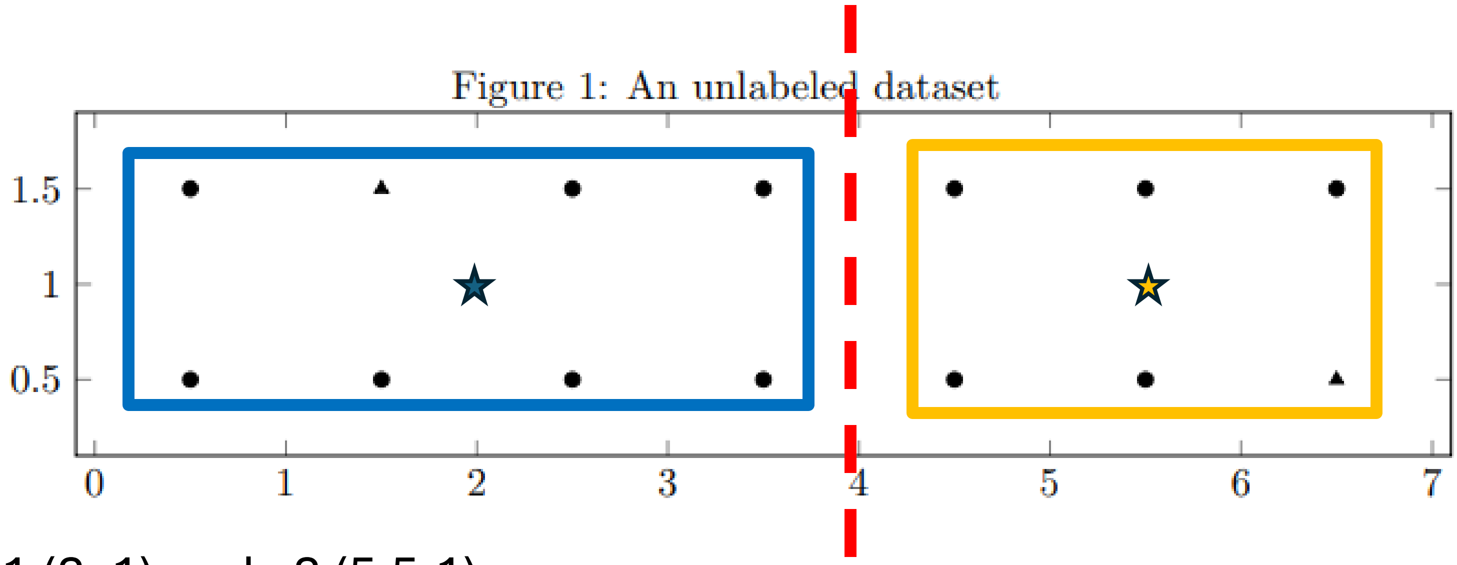
- New center: **(5.5,1)**

1 Clustering

Consider the following dataset. All points are unlabeled and part of the same set. The triangles are used to distinguish two points later.

Suppose that we run the K-means algorithm on this dataset with $K = 2$ and the two points indicated by triangles as the initial centers. When the algorithm converges, there will be two clearly separated clusters. Directly on Figure 1, draw a straight line that separates these two clusters, as well as the centers of these two clusters.

Figure 1: An unlabeled dataset



- Two new centers: c1 (2, 1) and c2 (5.5,1).
- Check for convergence:
    - **Check if there is any change of assignments of points.**
    - **There is none.**

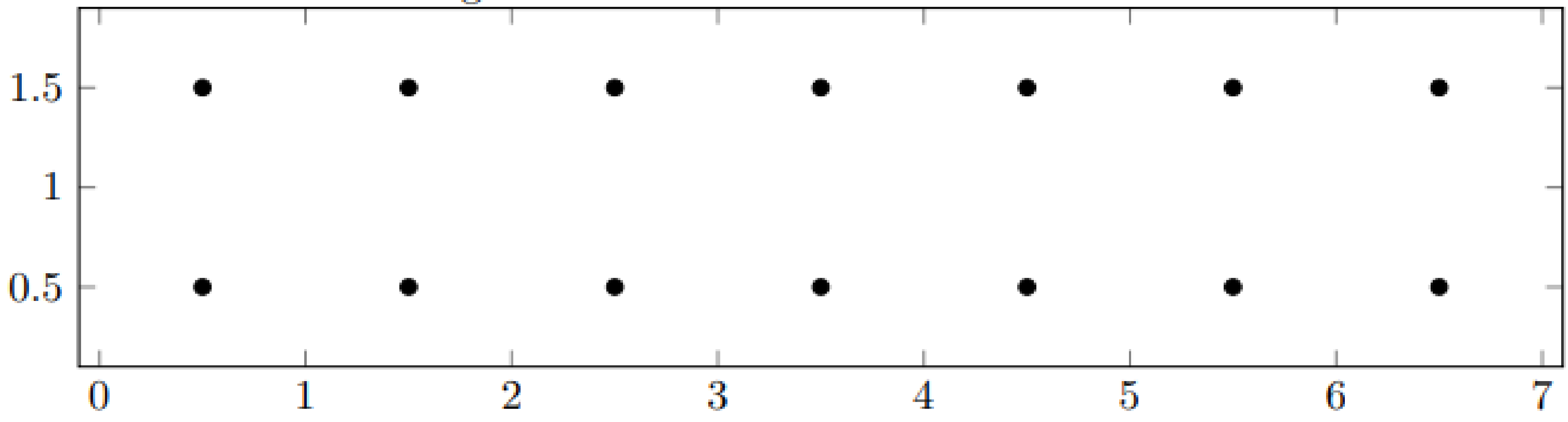
There are infinitely many choices for the separating line. Any vertical line between 3.5 and 4.5 would work for example.

The two centers are (2, 1) and (5.5, 1).

Next, find two other different sets of initialize centers that will converge to the exact same result if we apply K-means. Please follow the instructions below:

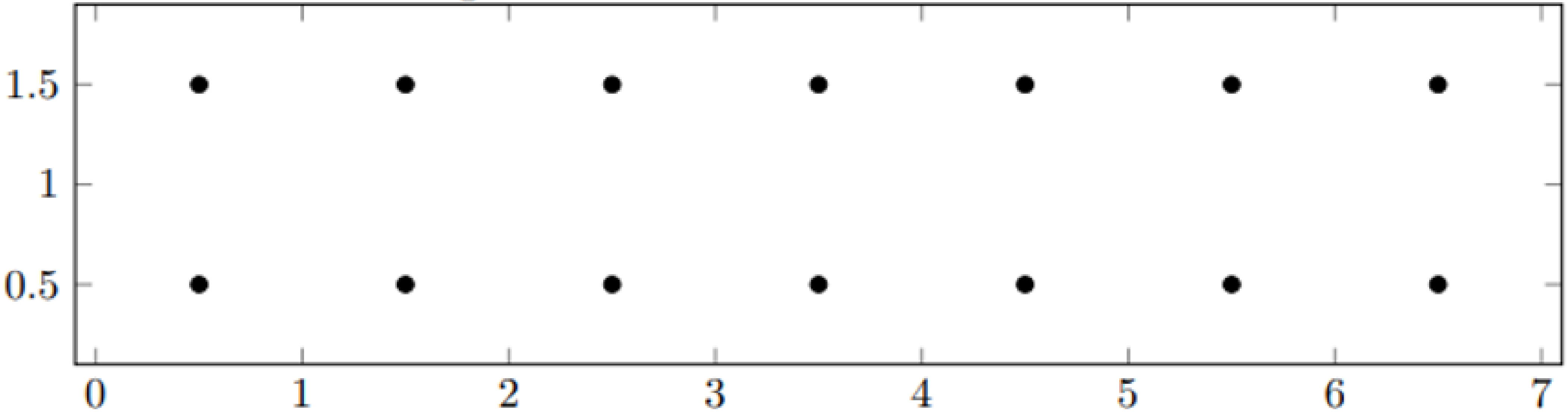
- the initialize centers have to be points of the dataset;
- directly on Figure 2, use two triangles to indicate the first set, and two squares to indicate the second;
- these two sets of points can overlap with each other, but of course cannot be the same;
- similarly these sets can overlap with the initialization of Figure 1 but cannot be the same;
- do not pick those that lead to ambiguous results due to different ways of breaking ties.

Figure 2: The same unlabeled dataset



Next, find two other different sets of initialize centers that will converge to the exact same result if we apply K-means. Please follow the instructions below:

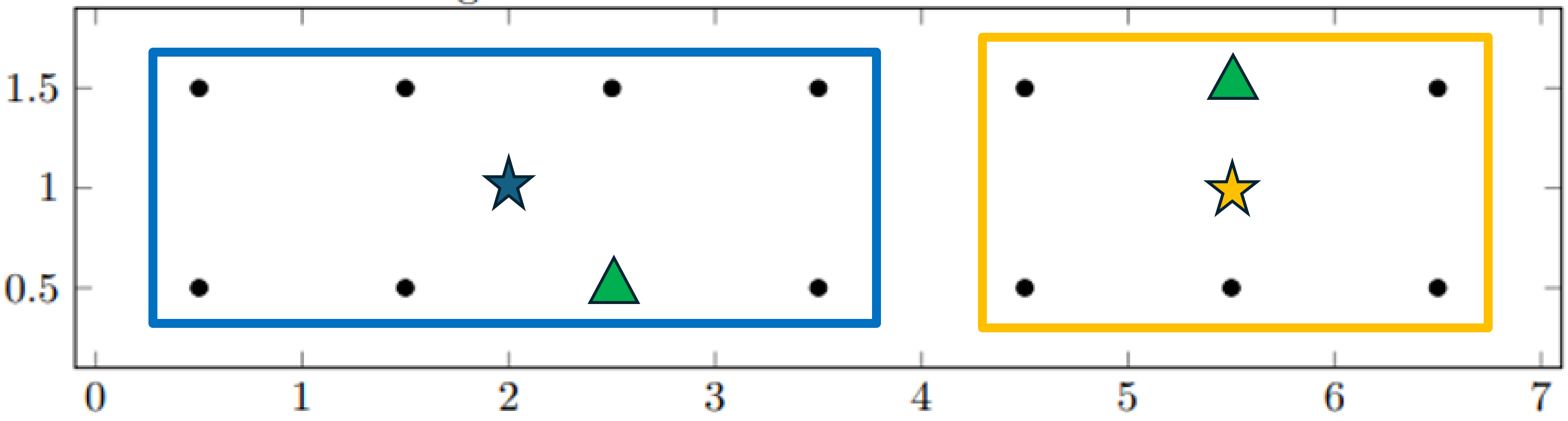
Figure 2: The same unlabeled dataset



There are many choices, for example, $(2.5, 0.5)$ and $(5.5, 1.5)$, $(2.5, 1.5)$ and $(5.5, 0.5)$, $(2.5, 1.5)$ and $(5.5, 1.5)$, $(1.5, 0.5)$ and $(6.5, 0.5)$, and so on. Note that answers like $(1.5, 0.5)$ and $(5.5, 0.5)$ are not acceptable due to the ambiguity from tie-breaking.

Next, find two other different sets of initialize centers that will converge to the exact same result if we apply K-means. Please follow the instructions below:

Figure 2: The same unlabeled dataset



- Two initial centers: c1 **(2.5, 0.5)** and c2 **(5.5,1.5).**
- Add points to cluster 1.
- Average x-coordinate: 2.
- Average y-coordinate: 1.

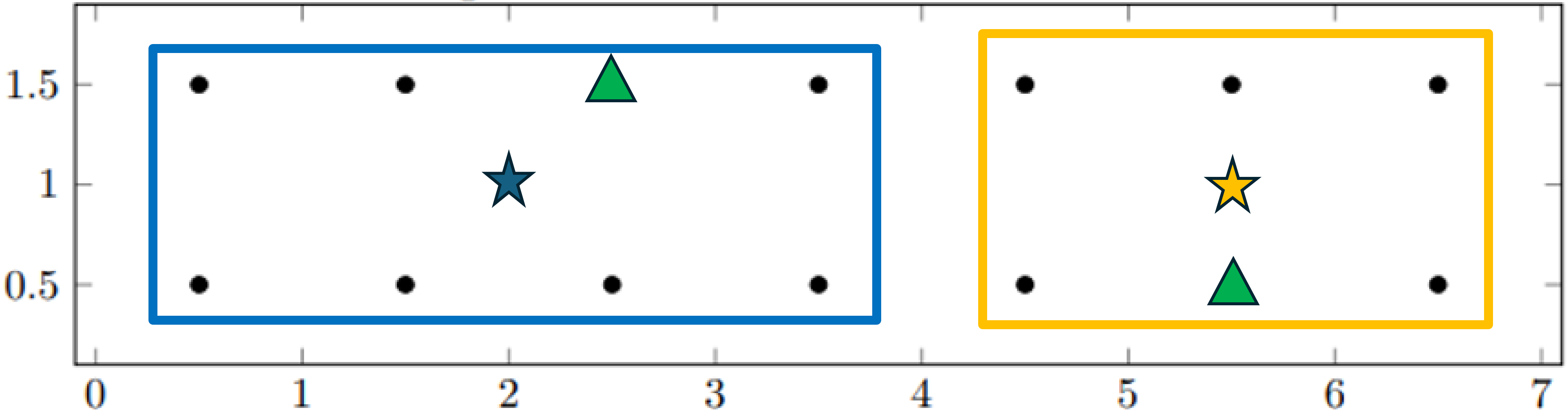
- New center: **(2,1).**

- Add points to cluster 2.
- Average x-coordinate: 5.5.
- Average y-coordinate: 1.

- New center: **(5.5,1).**

Next, find two other different sets of initialize centers that will converge to the exact same result if we apply K-means. Please follow the instructions below:

Figure 2: The same unlabeled dataset



- Two initial centers: c1 **(2.5, 1.5)** and c2 **(5.5,0.5).**
- Add points to cluster 1.
- Average x-coordinate: 2.
- Average y-coordinate: 1.

- New center: **(2,1).**

- Add points to cluster 2.
- Average x-coordinate: 5.5.
- Average y-coordinate: 1.

- New center: **(5.5,1).**

2 MLE, Mixtures, and EM

In this problem you will practice MLE and EM.

(a) Let $X \in \mathbb{R}$ be a random variable uniformly-distributed on some unknown interval $(0, \theta]$, where $\theta > 0$. More specifically, the density function is

$$
\begin{aligned}
P(X = x\,; \theta) &= \begin{cases} \frac{1}{\theta} & , \text{ if } x \in (0, \theta], \\ 0 & , \text{ otherwise}, \end{cases} && (1) \\
&= \frac{1}{\theta}\mathbf{1}[0 < x \leq \theta], && (2)
\end{aligned}
$$

where $\mathbf{1}[\cdot]$ is an indicator function that outputs 1 when the input condition is true, and 0 otherwise. Suppose that $x_1, x_2, \ldots, x_N$ are i.i.d. samples from this distribution. Write down the likelihood of the observations and then find the maximum likelihood estimator (MLE).

2 MLE, Mixtures, and EM

In this problem you will practice MLE and EM.

(a) Let $X \in \mathbb{R}$ be a random variable uniformly-distributed on some unknown interval $(0, \theta]$, where $\theta > 0$. More specifically, the density function is

$$P(X = x\,; \theta) = \begin{cases} \frac{1}{\theta} & \text{, if } x \in (0, \theta], \\ 0 & \text{, otherwise,} \end{cases} \tag{1}$$

$$= \frac{1}{\theta}\mathbf{1}[0 < x \leq \theta], \tag{2}$$

where $\mathbf{1}[\cdot]$ is an indicator function that outputs 1 when the input condition is true, and 0 otherwise. Suppose that $x_1, x_2, \ldots, x_N$ are i.i.d. samples from this distribution. Write down the likelihood of the observations and then find the maximum likelihood estimator (MLE).

- We know that the density function for "X=x" under θ is:

$$P(X = x\,; \theta) = \frac{1}{\theta}\mathbf{1}[0 < x \leq \theta],$$

- We observe data points $x_1, x_2, \ldots, x_N$, from this distribution.
- So the likelihood of seeing all these observations together is:

$$P(X = x_1; \theta) * P(X = x_2; \theta) * \cdots * P(X = x_n; \theta) = \quad P(x_1, \ldots, x_N\,; \theta) = \prod_{n=1}^{N} P(x_n\,; \theta)$$

2 MLE, Mixtures, and EM

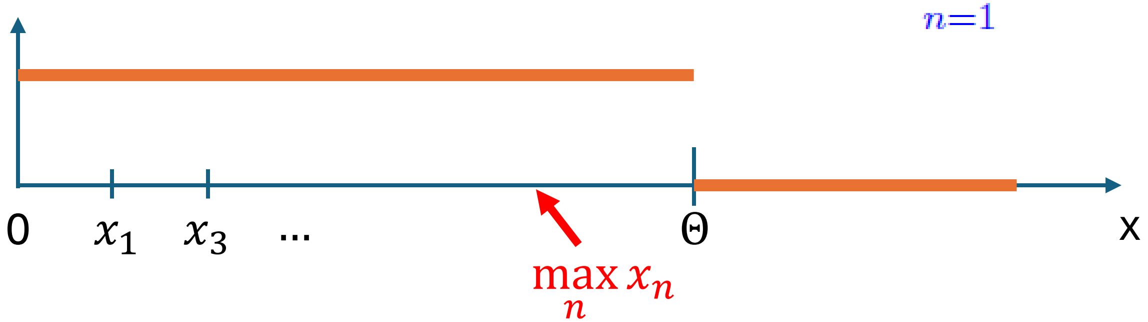
In this problem you will practice MLE and EM.

(a) Let $X \in \mathbb{R}$ be a random variable uniformly-distributed on some unknown interval $(0, \theta]$, where $\theta > 0$. More specifically, the density function is

$$
\begin{aligned}
P(X = x\,; \theta) &= \begin{cases} \frac{1}{\theta} & , \text{if } x \in (0, \theta], \\ 0 & , \text{otherwise}, \end{cases} && (1) \\
&= \frac{1}{\theta}\mathbf{1}[0 < x \le \theta], && (2)
\end{aligned}
$$

where $\mathbf{1}[\cdot]$ is an indicator function that outputs 1 when the input condition is true, and 0 otherwise. Suppose that $x_1, x_2, \dots, x_N$ are i.i.d. samples from this distribution. Write down the likelihood of the observations and then find the maximum likelihood estimator (MLE).

$$P(X = x_1; \theta) * \cdots * P(X = x_n; \theta) = P(x_1, \dots, x_N\,; \theta) = \prod_{n=1}^{N} P(x_n\,; \theta) = \frac{1}{\theta^N}\boxed{\mathbf{1}[\max_n x_n \le \theta]}$$



Why? Here is an example.

If the biggest value of x is ≤ θ, ($\max_n x_n \le \theta$), the product is 1 * 1 * ... * 1 = 1.

2 MLE, Mixtures, and EM

In this problem you will practice MLE and EM.

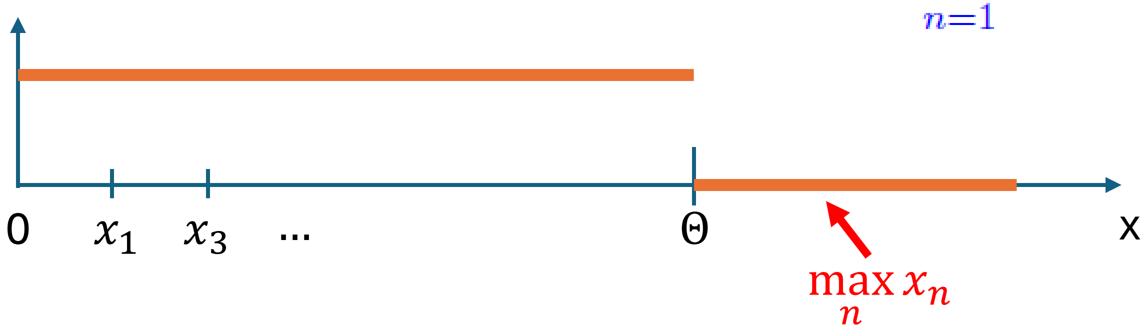
(a) Let $X \in \mathbb{R}$ be a random variable uniformly-distributed on some unknown interval $(0, \theta]$, where $\theta > 0$. More specifically, the density function is

$$P(X = x\,;\theta) = \begin{cases} \frac{1}{\theta} & , \text{if } x \in (0, \theta], \\ 0 & , \text{otherwise}, \end{cases} \tag{1}$$

$$= \frac{1}{\theta}\mathbf{1}[0 < x \le \theta], \tag{2}$$

where $\mathbf{1}[\cdot]$ is an indicator function that outputs 1 when the input condition is true, and 0 otherwise. Suppose that $x_1, x_2, \ldots, x_N$ are i.i.d. samples from this distribution. Write down the likelihood of the observations and then find the maximum likelihood estimator (MLE).

$$P(X = x_1; \theta) * \cdots * P(X = x_n; \theta) = P(x_1, ..., x_N\,;\theta) = \prod_{n=1}^{N} P(x_n\,;\theta) = \frac{1}{\theta^N}\boxed{\mathbf{1}[\max_n x_n \le \theta]}$$



Why? Here is an example.

If the biggest value of x is > θ, ($\max_n x_n > \theta$), the product is 1 * 1 * ... * **0** = 0.

2 MLE, Mixtures, and EM

In this problem you will practice MLE and EM.

(a) Let $X \in \mathbb{R}$ be a random variable uniformly-distributed on some unknown interval $(0, \theta]$, where $\theta > 0$. More specifically, the density function is

$$P(X = x\,; \theta) = \begin{cases} \frac{1}{\theta} & \text{, if } x \in (0, \theta], \\ 0 & \text{, otherwise,} \end{cases} \tag{1}$$

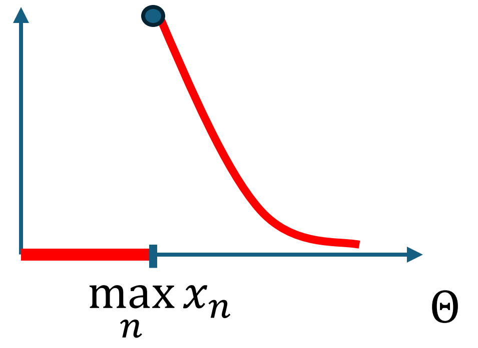
$$= \frac{1}{\theta}\mathbf{1}[0 < x \leq \theta], \tag{2}$$

where $\mathbf{1}[\cdot]$ is an indicator function that outputs 1 when the input condition is true, and 0 otherwise. Suppose that $x_1, x_2, \ldots, x_N$ are i.i.d. samples from this distribution. Write down the likelihood of the observations and then find the maximum likelihood estimator (MLE).

- Therefore, the likelihood is:

$$P(x_1, \ldots, x_N\,; \theta) = \prod_{n=1}^{N} P(x_n\,; \theta) = \frac{1}{\theta^N}\underline{\mathbf{1}[\max_n x_n \leq \theta]}$$

- To find the MLE:
  - If $\theta < \max_n x_n$, $P(x_1, \ldots, x_N; \theta) = 0$.
  - If $\theta \geq \max_n x_n$, $P(x_1, \ldots, x_N; \theta) = \frac{1}{\theta^N}$.
    - $\frac{1}{\theta^N}$ is a decreasing function of θ, given that N > 0.
    - So we have $\theta_* = \max_n x_n$



Roughly it looks like this, the shape of $\frac{1}{\theta^N}$ depends on the value of N.

Relevant materials: 10/24 slides page 35.

(b) Now suppose that X is distributed according to a **mixture** of two uniform distributions: one on interval $(0, \theta_1]$ and the other on $(0, \theta_2]$, for some unknown $\theta_1, \theta_2 > 0$. More specifically, the density function is

$$\begin{aligned} P(X = x) &= P(X = x, z = 1) + P(X = x, z = 2) \\ &= P(z = 1)P(X = x \mid z = 1) + P(z = 2)P(X = x \mid z = 2) \\ &= \omega_1 U(X = x\,; \theta_1) + \omega_2 U(X = x\,; \theta_2) \end{aligned}$$

where U is the uniform distribution defined as in Eq. (1) or Eq. (2), and ω_1, ω_2 are mixture weights such that

$$\omega_1 \geq 0, \omega_2 \geq 0, \text{ and } \omega_1 + \omega_2 = 1.$$

Suppose that $x_1, x_2, \ldots, x_N$ are i.i.d. samples from this mixture of uniform distributions. MLE does not admit a closed-form for this problem, and we will use the EM algorithm to approximately find the MLE.

- First, the E-Step fixes a set of parameters $\theta_1, \theta_2, \omega_1, \omega_2$ and computes for each n the posterior distribution $\gamma_{nk} = P(z_n = k \mid x_n\,; \theta_1, \theta_2, \omega_1, \omega_2)$ of the latent variable z_n, where $k \in \{1, 2\}$ indicates which mixture component x_n belongs to. Write down the explicit form of this posterior distribution *without using the proportional notation*. Then write down explicitly the expected complete log-likelihood using γ_{nk} (as a function of the four parameters $\theta_1, \theta_2, \omega_1, \omega_2$).

- X is distributed according to a **mixture** of two uniform distributions with unknown $\theta_1, \theta_2 > 0$.

$$\begin{aligned} P(X=x) &= P(X=x, z=1) + P(X=x, z=2) \\ &= P(z=1)P(X=x \mid z=1) + P(z=2)P(X=x \mid z=2) \\ &= \omega_1 U(X=x\,;\theta_1) + \omega_2 U(X=x\,;\theta_2) \end{aligned}$$

$$\omega_1 \geq 0, \omega_2 \geq 0, \text{ and } \omega_1 + \omega_2 = 1.$$

- U is the uniform distribution:

$$P(X=x\,;\theta) = \frac{1}{\theta}\mathbf{1}[0 < x \leq \theta],$$

- Suppose that $x_1, x_2, \ldots, x_N$ are i.i.d. samples from this mixture of uniform distributions.
- Use EM algorithm to approximately find the MLE.

First, the E-Step fixes a set of parameters $\theta_1, \theta_2, \omega_1, \omega_2$ and computes for each n the posterior distribution $\gamma_{nk} = P(z_n = k \mid x_n\,; \theta_1, \theta_2, \omega_1, \omega_2)$ of the latent variable z_n, where $k \in \{1, 2\}$ indicates which mixture component x_n belongs to.

- Write down the explicit form of this posterior distribution without using the proportional notation.
- write down explicitly the expected complete log-likelihood using γ_{nk}

- We do not know which distribution x_n is from.
- E-step: Estimate how likely each data point x_n comes from each component. $\gamma_{nk} = P(z_n = k \mid x_n\,; \theta_1, \theta_2, \omega_1, \omega_2)$

$$= \frac{P(z_n = k, x_n; \theta_1, \theta_2, w_1, w_2)}{P(x_n)}$$

By Bayes' rule

$$P(z_n = k \mid x_n\,; \theta_1, \theta_2, \omega_1, \omega_2) \propto P(z_n = k, x_n\,; \theta_1, \theta_2, \omega_1, \omega_2)$$

$$= P(z_n = k\,; \theta_1, \theta_2, \omega_1, \omega_2) P(x_n \mid z_n = k\,; \theta_1, \theta_2, \omega_1, \omega_2)$$

"is proportional to"

- X is distributed according to a **mixture** of two uniform distributions with unknown $\theta_1, \theta_2 > 0$.

$$\begin{aligned} P(X = x) &= P(X = x, z = 1) + P(X = x, z = 2) \\ &= P(z = 1)P(X = x \mid z = 1) + P(z = 2)P(X = x \mid z = 2) \\ &= \omega_1 U(X = x\,; \theta_1) + \omega_2 U(X = x\,; \theta_2) \end{aligned}$$

$$\omega_1 \geq 0, \omega_2 \geq 0, \text{ and } \omega_1 + \omega_2 = 1.$$

- U is the uniform distribution:

$$P(X = x\,; \theta) = \frac{1}{\theta}\mathbf{1}[0 < x \leq \theta],$$

- Suppose that $x_1, x_2, \ldots, x_N$ are i.i.d. samples from this mixture of uniform distributions.
- Use EM algorithm to approximately find the MLE.

First, the E-Step fixes a set of parameters $\theta_1, \theta_2, \omega_1, \omega_2$ and computes for each n the posterior distribution $\gamma_{nk} = P(z_n = k \mid x_n\,; \theta_1, \theta_2, \omega_1, \omega_2)$ of the latent variable z_n, where $k \in \{1, 2\}$ indicates which mixture component x_n belongs to.

- Write down the explicit form of this posterior distribution without using the proportional notation.
- write down explicitly the expected complete log-likelihood using γ_{nk}

$$= P(z_n = k\,; \theta_1, \theta_2, \omega_1, \omega_2)P(x_n \mid z_n = k\,; \theta_1, \theta_2, \omega_1, \omega_2)$$

From the problem definition: $P(z_n = k) = \omega_k,$ $\quad P(x_n \mid z_n = k) = \frac{1}{\theta_k}\mathbf{1}[0 < x_n \leq \theta_k].$

$$= \omega_k U(X = x_n\,; \theta_k)$$

Put in the definitions.

$$= \frac{\omega_k}{\theta_k}\mathbf{1}[0 < x_n \leq \theta_k]$$

- X is distributed according to a **mixture** of two uniform distributions with unknown $\theta_1, \theta_2 > 0$.

$$\begin{aligned} P(X = x) &= P(X = x, z = 1) + P(X = x, z = 2) \\ &= P(z = 1)P(X = x \mid z = 1) + P(z = 2)P(X = x \mid z = 2) \\ &= \omega_1 U(X = x \,; \theta_1) + \omega_2 U(X = x \,; \theta_2) \end{aligned}$$

$$\omega_1 \geq 0, \omega_2 \geq 0, \text{ and } \omega_1 + \omega_2 = 1.$$

- U is the uniform distribution:

$$P(X = x \,; \theta) = \frac{1}{\theta}\mathbf{1}[0 < x \leq \theta],$$

- Suppose that $x_1, x_2, \ldots, x_N$ are i.i.d. samples from this mixture of uniform distributions.
- Use EM algorithm to approximately find the MLE.

  First, the E-Step fixes a set of parameters $\theta_1, \theta_2, \omega_1, \omega_2$ and computes for each n the posterior distribution $\gamma_{nk} = P(z_n = k \mid x_n \,; \theta_1, \theta_2, \omega_1, \omega_2)$ of the latent variable z_n, where $k \in \{1, 2\}$ indicates which mixture component x_n belongs to.

- Write down the explicit form of this posterior distribution without using the proportional notation.
- write down explicitly the expected complete log-likelihood using γ_{nk}

$$= \frac{P(z_n = k, x_n; \theta_1, \theta_2, w_1, w_2)}{P(x_n)} \checkmark$$

We have shown that, $P(z_n = k, x_n; \theta_1, \theta_2, w_1, w_2) = \frac{\omega_k}{\theta_k}\mathbf{1}[0 < x_n \leq \theta_k]$

$$P(x_n) = \omega_1 \frac{1}{\theta_1}\mathbf{1}[0 < x_n \leq \theta_1] + \omega_2 \frac{1}{\theta_2}\mathbf{1}[0 < x_n \leq \theta_2].$$

Marginal likelihood.

Thus,

$$\gamma_{nk} = P(z_n = k \mid x_n \,; \theta_1, \theta_2, \omega_1, \omega_2) = \frac{\frac{\omega_k}{\theta_k}\mathbf{1}[0 < x_n \leq \theta_k]}{\frac{\omega_1}{\theta_1}\mathbf{1}[0 < x_n \leq \theta_1] + \frac{\omega_2}{\theta_2}\mathbf{1}[0 < x_n \leq \theta_2]}, \;\forall\, k \in \{1, 2\}.$$

- X is distributed according to a **mixture** of two uniform distributions with unknown $\theta_1, \theta_2 > 0$.

$$\begin{aligned} P(X = x) &= P(X = x, z = 1) + P(X = x, z = 2) \\ &= P(z = 1) P(X = x \mid z = 1) + P(z = 2) P(X = x \mid z = 2) \\ &= \omega_1 U(X = x \,; \theta_1) + \omega_2 U(X = x \,; \theta_2) \end{aligned}$$

$$\omega_1 \geq 0, \omega_2 \geq 0, \text{ and } \omega_1 + \omega_2 = 1.$$

- U is the uniform distribution:

$$P(X = x \,; \theta) = \frac{1}{\theta} \mathbf{1}[0 < x \leq \theta],$$

- Suppose that $x_1, x_2, \ldots, x_N$ are i.i.d. samples from this mixture of uniform distributions.
- Use EM algorithm to approximately find the MLE.

First, the E-Step fixes a set of parameters $\theta_1, \theta_2, \omega_1, \omega_2$ and computes for each n the posterior distribution $\gamma_{nk} = P(z_n = k \mid x_n \,; \theta_1, \theta_2, \omega_1, \omega_2)$ of the latent variable z_n, where $k \in \{1, 2\}$ indicates which mixture component x_n belongs to.

- Write down the explicit form of this posterior distribution without using the proportional notation.
- write down explicitly the expected complete log-likelihood using γ_{nk}

---

and obtain **Expectation** of complete likelihood

$$Q(\boldsymbol{\theta} \,; \boldsymbol{\theta}^{(t)}) = \sum_{n=1}^{N} \mathbb{E}_{z_n \sim q_n^{(t)}} [\ln p(\boldsymbol{x}_n, z_n \,; \boldsymbol{\theta})]$$

$$\ln P(x_n, z_n = k \,; \theta_1, \theta_2, \omega_1, \omega_2)$$

The expectation over the distribution of the latent variable z_n

$$Q(\theta_1, \theta_2, \omega_1, \omega_2) = \sum_n \sum_k \gamma_{nk} \ln P(x_n, z_n = k \,; \theta_1, \theta_2, \omega_1, \omega_2)$$

$$P(z_n = k, x_n; \theta_1, \theta_2, w_1, w_2) = \frac{\omega_k}{\theta_k} 1[0 < x_n \leq \theta_k]$$

$$= \sum_n \sum_k \gamma_{nk} \ln \omega_k + \sum_n \sum_k \gamma_{nk} \ln \frac{1[0 < x_n \leq \theta_k]}{\theta_k}.$$

$$\ln(ab) = \ln a + \ln b.$$

- X is distributed according to a **mixture** of two uniform distributions with unknown $\theta_1, \theta_2 > 0$.

$$\begin{aligned} P(X=x) &= P(X=x, z=1) + P(X=x, z=2) \\ &= P(z=1)P(X=x \mid z=1) + P(z=2)P(X=x \mid z=2) \\ &= \omega_1 U(X=x\,;\theta_1) + \omega_2 U(X=x\,;\theta_2) \end{aligned}$$

$$\omega_1 \geq 0, \omega_2 \geq 0, \text{ and } \omega_1 + \omega_2 = 1.$$

- U is the uniform distribution:

$$P(X=x\,;\theta) = \frac{1}{\theta}\mathbf{1}[0 < x \leq \theta],$$

- Suppose that $x_1, x_2, \ldots, x_N$ are i.i.d. samples from this mixture of uniform distributions.
- Use EM algorithm to approximately find the MLE.
  - Next, derive the M-Step by maximizing the expected complete log-likelihood you derived from the last problem over the four parameters. Hint: you will need to use the fact $0 \ln 0 = 0$.

Step 2 (M-Step) update the model parameter via **Maximization**

$$\boldsymbol{\theta}^{(t+1)} \leftarrow \operatorname*{argmax}_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}\,;\boldsymbol{\theta}^{(t)})$$

- M-step: update our guesses for $\theta_1, \theta_2, w_1, w_2$

$$Q(\theta_1, \theta_2, \omega_1, \omega_2) = \underline{\sum_n \sum_k \gamma_{nk} \ln \omega_k} + \sum_n \sum_k \gamma_{nk} \ln\left(\frac{\mathbf{1}[0 < x_n \leq \theta_k]}{\theta_k}\right)$$

- Solve for w_k (10/24 lecture slides page 47 and 48)

To find $\omega_1, \ldots, \omega_K$, solve

$$\operatorname*{argmax}_{\boldsymbol{\omega}} \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \ln \omega_k \qquad\qquad \omega_k = \frac{\sum_n \gamma_{nk}}{N}$$

- X is distributed according to a **mixture** of two uniform distributions with unknown $\theta_1, \theta_2 > 0$.

$$\begin{aligned} P(X=x) &= P(X=x, z=1) + P(X=x, z=2) \\ &= P(z=1)P(X=x \mid z=1) + P(z=2)P(X=x \mid z=2) \\ &= \omega_1 U(X=x\,;\theta_1) + \omega_2 U(X=x\,;\theta_2) \end{aligned}$$

$$\omega_1 \geq 0, \omega_2 \geq 0, \text{ and } \omega_1 + \omega_2 = 1.$$

- U is the uniform distribution:

$$P(X=x\,;\theta) = \frac{1}{\theta}\mathbf{1}[0 < x \leq \theta],$$

- Suppose that $x_1, x_2, \ldots, x_N$ are i.i.d. samples from this mixture of uniform distributions.
- Use EM algorithm to approximately find the MLE.
  - Next, derive the M-Step by maximizing the expected complete log-likelihood you derived from the last problem over the four parameters. Hint: you will need to use the fact $0 \ln 0 = 0$.

Step 2 (M-Step) update the model parameter via Maximization

$$\boldsymbol{\theta}^{(t+1)} \leftarrow \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}\,;\boldsymbol{\theta}^{(t)})$$

- M-step: update our guesses for $\theta_1, \theta_2, w_1, w_2$

$$Q(\theta_1, \theta_2, \omega_1, \omega_2) = \sum_n \sum_k \gamma_{nk} \ln \omega_k + \underline{\sum_n \sum_k \gamma_{nk} \ln\left(\frac{\mathbf{1}[0 < x_n \leq \theta_k]}{\theta_k}\right)}$$

- Solve for θ_k:

$$\arg\max_{\theta_k} \sum_n \gamma_{nk} \ln \frac{\mathbf{1}[0 < x_n \leq \theta_k]}{\theta_k} = \arg\max_{\theta_k} \sum_{n:\gamma_{nk}>0} \gamma_{nk} \ln \frac{\mathbf{1}[0 < x_n \leq \theta_k]}{\theta_k}$$

Drop points with $\gamma_{nk} = 0$

- X is distributed according to a **mixture** of two uniform distributions with unknown $\theta_1, \theta_2 > 0$.

$$\begin{aligned} P(X=x) &= P(X=x, z=1) + P(X=x, z=2) \\ &= P(z=1)P(X=x \mid z=1) + P(z=2)P(X=x \mid z=2) \\ &= \omega_1 U(X=x\,;\theta_1) + \omega_2 U(X=x\,;\theta_2) \end{aligned}$$

$$\omega_1 \geq 0, \omega_2 \geq 0, \text{ and } \omega_1 + \omega_2 = 1.$$

- U is the uniform distribution:

$$P(X=x\,;\theta) = \frac{1}{\theta}\mathbf{1}[0 < x \leq \theta],$$

- Suppose that $x_1, x_2, \ldots, x_N$ are i.i.d. samples from this mixture of uniform distributions.
- Use EM algorithm to approximately find the MLE.
    - Next, derive the M-Step by maximizing the expected complete log-likelihood you derived from the last problem over the four parameters. Hint: you will need to use the fact $0 \ln 0 = 0$.

- M-step: update our guesses for $\theta_1, \theta_2, w_1, w_2$

$$Q(\theta_1, \theta_2, \omega_1, \omega_2) = \sum_n \sum_k \gamma_{nk} \ln \omega_k + \underline{\sum_n \sum_k \gamma_{nk} \ln \left(\frac{\mathbf{1}[0 < x_n \leq \theta_k]}{\theta_k} \right)}$$

- Solve for θ_k:

Similar to Problem 2 (a) in this exercise.

$$\arg\max_{\theta_k} \sum_{n:\gamma_{nk}>0} \gamma_{nk} \ln \frac{\mathbf{1}[0 < x_n \leq \theta_k]}{\theta_k}$$

- If $\theta_k < \max_{n:\gamma_{nk}>0} x_n$, ln(-) function goes to -∞.
- If $\theta_k \geq \max_{n:\gamma_{nk}>0} x_n$, $\ln \frac{\mathbf{1}[0 < x_n \leq \theta_k]}{\theta_k}$ is a decreasing function of θ_k
- So, the max is: $\max_{n:\gamma_{nk}>0} x_n$