Project 1: Color Segmentation Solutions

Problems

In square brackets are the points assigned to each part.

1. [26 pts] Let $U_1, \dots, U_n$ be independent, identically distributed *Uniform* random variables with (continuous) support on $(0, b)$, where $b > 0$ is a parameter.

   (a) Define the random variable $Y := -\sum_{i=1}^{n} \log(U_i)$, where log is the natural logarithm function. Determine the probability density function (pdf) $p(y; b)$ of Y by explicitly computing it.

   (b) Based on the pdf you found in part (a) above, determine the third moment of Y, i.e., $\mathbb{E}[Y^3]$.

   (c) Suppose now that you are given two independent observations y_1 and y_2 of Y. Determine the maximum likelihood estimate of b based on the observations y_1 and y_2.

   Solution

   (a) [9 pts] Let $X_i = f(U_i) := -\log U_i$ and $p_{x_i}(x) = \frac{1}{|\det\left(\frac{df}{du}(f^{-1}(x))\right)|} p_{u_i}(f^{-1}(x))$

$$\frac{df}{du} = \frac{-1}{u}, \quad f^{-1}(x) = e^{-x} \Rightarrow p_{x_i}(x; b) = \frac{e^{-x}}{b}$$

$$U_i \in (0, b) \Rightarrow X_i \in (-\log b, \infty)$$

$$\Rightarrow X_i \sim p_x(x; b) = \frac{e^{-x}}{b} \in (-\log b, \infty)$$

Let $Y_1 = X_1$,

$$Y_{i+1} := Y_i + X_{i+1} \iff [p_{y_i} * p_x](y) = \int p_{y_i}(y - x) p_x(x) dx$$

For $i = 2$,

$$x \in (-\log b, \infty), \quad y - x \in (-\log b, \infty) \Rightarrow x \in (-\log b, \log b + y), \quad y \in (-2\log b, \infty)$$

$$p_{y_2}(y; b) = \int_{-\log b}^{\log b + y} \frac{e^{x-y}}{b} \frac{e^{-x}}{b} dx = \frac{e^{-y}}{b^2} x \Big|_{-\log b}^{\log b + y} = \frac{e^{-y}}{b^2}(2\log b + y) \in (-2\log b, \infty)$$

$$p_{y_3}(y; b) = [p_{y_2} * p_{x_3}](y) = \frac{e^{-y}}{2b^3}(3\log b + y)^2 \in (-3\log b, \infty)$$

Assume $p_{y_n}(y; b) = p_y(y; b) = \frac{e^{-y}}{(n-1)! b^n}(y + n\log b)^{n-1}$ and let $Y_{n+1} = Y + X_{n+1}$

$$p_{y_{n+1}}(y; b) = \int_{-n\log b}^{\log b + y} \frac{e^{x-y}}{b} \frac{e^{-x}}{(n-1)! b^n}(n\log b + x)^{n-1} dx = \frac{e^{-y}}{(n-1)! b^{n+1}} \int_{-n\log b}^{\log b + y} (n\log b + x)^{n-1} dx$$

$$= \frac{e^{-y}}{(n-1)! b^{n+1}} \frac{(n\log b + x)^n}{n} \Big|_{-n\log b}^{\log b + y} = \frac{e^{-y}}{n! b^{n+1}}(y + (n+1)\log b)^n$$

$$\Rightarrow Y \sim p_y(y; b) = \frac{e^{-y}}{(n-1)! b^n}(y + n\log b)^{n-1} \qquad Y \in (-n\log b, \infty)$$

   (b) [8 pts] $Y \in (-n\log b, \infty)$ Let $Z = Y + n\log b \in (0, +\infty)$ so $Y = Z - n\log b$. Then

$$q(z) = p(y; b) = \frac{e^{-y}}{(n-1)! b^n}(y + n\log b)^{n-1} = \frac{e^{-(z - n\log b)}}{(n-1)! b^n} z^{n-1} = \frac{e^{-z} z^{n-1}}{(n-1)!}$$

We find that Z is a Gamma distribution: $Z \sim \text{Gamma}(\alpha, \beta) = \text{Gamma}(n, 1)$. And use the property of Gamma distribution:

$$\mathbb{E}[Z^k] = \frac{(\alpha + k - 1)!}{(\alpha - 1)! \beta^k} = \frac{(n + k - 1)!}{(n - 1)!}$$

$$\begin{aligned}
\mathbb{E}[Z^k] &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-\beta z} z^{\alpha+k-1} dz \\
&= \frac{1}{\Gamma(\alpha)\beta^k} \int_0^\infty e^{-t} t^{\alpha+k-1} dt \\
&= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\beta^k} \\
&= \frac{(n+k-1)!}{(n-1)!}
\end{aligned}$$

Let $\Gamma(n) := \int_0^\infty e^{-t} t^{n-1} dt$. If $n \in \mathbb{Z}$ then $\Gamma(n) = (n-1)!$

$$\begin{aligned}
\Gamma(n) &= \int_0^\infty e^{-t} t^{n-1} dt \\
&= t^{n-1} e^{-t} \Big|_0^\infty + (n-1) \int_0^\infty e^{-t} t^{n-2} dt \\
&= \lim_{p\to\infty} \frac{t^{n-1}}{e^t} \Big|_0^p + (n-1) \int_0^\infty e^{-t} t^{n-2} dt \\
&= (n-1) \int_0^\infty e^{-t} t^{n-2} dt \\
&= \dots \\
&= (n-1)(n-2)(n-3)\dots(2)(1) \\
&= (n-1)!
\end{aligned}$$

So we have $\mathbb{E}[Z] = n, \mathbb{E}[Z^2] = n(n+1), \mathbb{E}[Z^3] = n(n+1)(n+2)$

$$\begin{aligned}
\mathbb{E}[Y^3] &= \mathbb{E}[(Z - n\log b)^3] \\
&= \mathbb{E}[Z^3] - 3n\log b\mathbb{E}[Z^2] + 3(n\log b)^2\mathbb{E}[Z] - (n\log b)^3 \\
&= n(n+1)(n+2) - 3\log(b)n^2(n+1) + 3\log^2(b)n^3 - \log^3(b)n^3 \\
&= n(n+1)(n+2) - 3\log(b)n^2(n+1) + \log^2(b)(3 - \log(b))n^3
\end{aligned}$$

(Another method)
$Y \in (-n\log b, \infty)$ Let $Z = Y + n\log b \in (0, +\infty)$ so $Y = Z - n\log b$. Then

$$\begin{aligned}
\mathbb{E}[Y^3] &= \int_{-n\log b}^\infty \frac{e^{-y}}{(n-1)!b^n} y^3 (y + n\log b)^{n-1} dy \\
&= \frac{1}{(n-1)!} \int_0^\infty (z - n\log b)^3 e^{-z} z^{n-1} dz \\
&= \frac{1}{(n-1)!} \left[(n+2)! - 3n\log b(n+1)! + 3n^2\log^2 b(n)! + n^3\log^3 b(n-1)!\right] \\
&= (n+2)(n+1)n - 3n^2(n+1)\log b + 3n^3\log b + n^3\log^3 b
\end{aligned}$$

Show $\Gamma(n) = (n-1)!$ if $n \in \mathbb{Z}$ for completeness.

(c) [9 pts] To maximize $p(y_1, y_2|b)$, we should have $\min(y_1, y_2) > -n\log b \Leftrightarrow b > e^{-\frac{1}{n}\min(y_1,y_2)}$.

In (b) we define $Z = Y + n\log b$, so here we can let $z_1 = y_1 + n\log b, z_2 = y_2 + n\log b$, then

$$p(y_1, y_2|b) = q(z_1, z_2|b) = \frac{e^{-(z_1+z_2)}(z_1z_2)^{n-1}}{((n-1)!)^2}$$

$$\log q(z_1, z_2|b) = -(z_1 + z_2) + (n-1)(\log z_1 + \log z_2) - 2\log((n-1)!)$$
$$\log p(y_1, y_2|b) = -(y_1 + y_2 + 2n\log b) + (n-1)(\log(y_1 + n\log b) + \log(y_2 + n\log b)) - 2\log((n-1)!)$$
$$\frac{\partial \log p}{\partial b} = \frac{-2n}{b} + (n-1)\left(\frac{n}{b(y_1 + n\log b)} + \frac{n}{b(y_2 + n\log b)}\right) = 0$$
$$\left(\frac{1}{y_1 + n\log b} + \frac{1}{y_2 + n\log b}\right) = \frac{2}{n-1}$$

Replace $\log b$ with c and rewrite as quadratic below.

(Another method)

$$\begin{aligned} p(y_1, y_2|b) &= \frac{e^{-y_1}}{(n-1)!b^n}(y_1 + n\log b)^{n-1}\frac{e^{-y_2}}{(n-1)!b^n}(y_2 + n\log b)^{n-1} \\ &= \frac{e^{-(y_1+y_2)}}{((n-1)!)^2 b^{2n}}[(y_1 + n\log b)(y_2 + n\log b)]^{n-1} \\ &= \frac{e^{-(y_1+y_2+2n\log b)}}{((n-1)!)^2}[(y_1 + n\log b)(y_2 + n\log b)]^{n-1} \end{aligned}$$

Take the derivative w.r.t. b and discard terms without b we get

$$-2(y_1 + n\log b)(y_2 + n\log b) + (n-1)((y_1 + y_2 + 2n\log b)) = 0$$

Replace $\log b$ with c:

$$-2(y_1 + nc)(y_2 + nc) + (n-1)(y_1 + y_2 + 2nc) = 0$$
$$-2n^2c^2 - 2n(y_1 + y_2)c - 2y_1y_2 + 2(n-1)nc + (n-1)(y_1 + y_2) = 0$$
$$2n^2c^2 + 2n(y_1 + y_2 + 1 - n)c + 2y_1y_2 - (n-1)(y_1 + y_2) = 0$$
$$c = \frac{1}{2n}\left[-(y_1 + y_2 + 1 - n) \pm \sqrt{(y_1 + y_2 + 1 - n)^2 - 4y_1y_2 + 2(n-1)(y_1 + y_2)}\right]$$
$$c = \frac{1}{2n}\left[-(y_1 + y_2 + 1 - n) \pm \sqrt{(y_1 - y_2)^2 + (n-1)^2}\right]$$

Choose b^*_{MLE} based on $b > e^{-\frac{1}{n}\min(y_1, y_2)}$ and $b = e^{\frac{1}{2n}\left(-(y_1+y_2+1-n)\pm\sqrt{(y_1-y_2)^2+(n-1)^2}\right)}$

2. [26 pts] Consider the Logistic Regression (LR) and Gaussian Naive Bayes (GNB) models in the K-class setting with D features. As a reminder, the generative and discriminative models used by LR and GNB, respectively, for a given labeled example $(\mathbf{x}, y)$ with $\mathbf{x} \in \mathbb{R}^D$ and $y = k \in \{1, \ldots, K\}$ are:

$$\text{LR}: \qquad p(y \mid \mathbf{x}) = \frac{\exp\left(\sum_{d=1}^{D} \omega_{d,k} x_d\right)}{\sum_{j=1}^{K} \exp\left(\sum_{d=1}^{D} \omega_{d,j} x_d\right)}$$

$$\text{GNB}: \qquad p(y, \mathbf{x}) = \theta_k \prod_{d=1}^{D} \phi(x_d; \mu_{d,k}, \sigma_{d,k}^2)$$

(a) How many parameters must be estimated by:

- Logistic Regression
- Logistic Regression with additional quadratic features of the form $x_i x_j$ for all $i, j \in \{1, \ldots, D\}$?
- Gaussian Naive Bayes with arbitrary variances
- Gaussian Naive Bayes with variance shared among the classes, i.e., the variance of feature d for class k is $\sigma_{d,k}^2 = \sigma_d^2$ (independent of k)?

(b) In this part, we will derive the gradient ascent algorithm for optimizing the weights $\boldsymbol{\omega}_k := [\omega_{1,k}, \; \cdots \; , \omega_{D,k}]^T \in \mathbb{R}^D$ for $k = 1, \ldots, K$ of K-ary Logistic Regression.

- Explicitly write down the log-likelihood as a function of the parameters, $J(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_K)$.
- Note that there is no closed-form solution to $\max_{\boldsymbol{\omega}_k} J(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_K)$ but we can still find a solution using gradient ascent. Derive an expression for the k-th component of the gradient of $J(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_K)$ with respect to $\boldsymbol{\omega}_k$.
- Beginning with initial weights $\boldsymbol{\omega}_k^{(t)}$, write down the update rule for $\boldsymbol{\omega}_k^{(t)}$ using gradient ascent with step size α.

Solution

(a) [3 pts each]

- Logistic Regression: KD
- Logistic Regression with quadratic features: $K(D+D(D-1)/2+D)$ because there are $D(D-1)/2$ pairs of features of the form $x_i x_j$ and D features of the form $x_i x_i$.
- Gaussian Naive Bayes: $K(D + D) + K = K(2D + 1)$
- Gaussian Naive Bayes with shared variance: $KD + D + K$

(b) [5,5,4 pts]

- $\log p(y \mid \mathbf{x}) = \log\left(\frac{\exp(\boldsymbol{\omega}_y^T \mathbf{x})}{\sum_{i=1}^{K} \exp(\boldsymbol{\omega}_i^T \mathbf{x})}\right) = \boldsymbol{\omega}_y^T \mathbf{x} - \log\left(\sum_{i=1}^{K} \exp(\boldsymbol{\omega}_i^T \mathbf{x})\right)$

  $\log p(y_{1:n} \mid \mathbf{x}_{1:n}) = \sum_{i=1}^{n} \log\left(\frac{\exp(\boldsymbol{\omega}_y^T \mathbf{x}_i)}{\sum_{i=1}^{K} \exp(\boldsymbol{\omega}_i^T \mathbf{x}_i)}\right) = \sum_{i=1}^{n} \left(\boldsymbol{\omega}_y^T \mathbf{x}_i - \log\left(\sum_{i=1}^{K} \exp(\boldsymbol{\omega}_i^T \mathbf{x}_i)\right)\right)$
- $\nabla_{\boldsymbol{\omega}_k} J(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_K) = \mathbf{x} \mathbf{1}_{\{y=k\}} - \frac{\exp(\boldsymbol{\omega}_k^T \mathbf{x})}{\sum_{i=1}^{K} \exp(\boldsymbol{\omega}_i^T \mathbf{x})} \mathbf{x} = (\mathbf{1}_{\{y=k\}} - p(y = k|\mathbf{x}))\mathbf{x} = (\mathbf{1}_{\{y=k\}} - \exp(J(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_K)))\mathbf{x}$
- $\boldsymbol{\omega}_k^{(t+1)} = \boldsymbol{\omega}_k^{(t)} + \alpha \nabla_{\boldsymbol{\omega}_k} J(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_K) = \boldsymbol{\omega}_k^{(t)} + \alpha(\mathbf{1}_{\{y=k\}} - \exp(J(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_K)))\mathbf{x}$

3. [26 pts] Consider a data set $\mathcal{D} = \left\{ \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 5 \\ * \end{pmatrix}, \begin{pmatrix} * \\ 6 \end{pmatrix} \right\}$, where the $*$ symbol indicates missing values. Suppose that the data was generated using independent draws from a two-dimensional *Uniform* distribution:

$$p(\mathbf{x}; \boldsymbol{\theta}) = \begin{cases} \frac{1}{|u_1 - l_1||u_2 - l_2|}, & \text{if } l_1 \le x_1 \le u_1 \text{ and } l_2 \le x_2 \le u_2 \\ 0, & \text{otherwise,} \end{cases}$$

where $\boldsymbol{\theta} := \begin{bmatrix} l_1 & u_1 & l_2 & u_2 \end{bmatrix}^T$ are the distribution parameters. Our goal is to determine the maximum likelihood estimate of $\boldsymbol{\theta}$, based on the known data values $\mathcal{D}_{known}$, by solving:

$$\max_{\theta} \log p(\mathcal{D}_{known}; \boldsymbol{\theta})$$

Since there is missing data, we will use the EM algorithm to estimate $\boldsymbol{\theta}$ by thinking of the missing values as latent variables.

(a) (E step) Starting with an initial estimate $\boldsymbol{\theta}^{(0)} := \begin{bmatrix} 0 & 10 & 0 & 10 \end{bmatrix}^T$, explicitly compute an upper bound $\mathcal{T}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(0)})$ to the data log-likelihood.

(b) (M step) Find $\boldsymbol{\theta}^{(1)}$ that maximizes $\mathcal{T}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(0)})$.

(c) Make a 2-D plot of the data and the bounding box determined by your parameter estimate $\boldsymbol{\theta}^{(1)}$.

(d) Determine the final MLE estimate $\boldsymbol{\theta}_{MLE}$ that the EM algorithm will eventually converge to.

Solution

(a) [10 pts] Let $\theta^{(0)} = (0, 10, 0, 10)^t$
E-Step:

$$\begin{aligned}
\tau(\theta, \theta^0) &= \mathbb{E}[\ln p(D_{known}; \theta \mid \theta^0)] \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\sum_{k=1}^{3} \ln p(x_k \mid \theta) + \ln p(x_4 \mid \theta) + \ln p(x_5 \mid \theta)] p(x_{42} \mid \theta_0; x_{41} = 5) p(x_{51} \mid \theta_0; x_{52} = 6) dx_{42} dx_{51} \\
&= \sum_{k=1}^{3} \ln p(x_k \mid \theta) + \int_{-\infty}^{\infty} \ln p(x_4 \mid \theta) p(x_{42} \mid \theta_0; x_{41} = 5) dx_{42} + \int_{-\infty}^{\infty} \ln p(x_5 \mid \theta) p(x_{51} \mid \theta_0; x_{52} = 6) dx_{51}
\end{aligned}$$

Let the second term be K and third term be L:

$$\begin{aligned}
K &= \int_{-\infty}^{\infty} \ln p(x_4 \mid \theta) p(x_{42} \mid \theta_0; x_{41} = 5) dx_{42} \\
&= \int_{-\infty}^{\infty} \ln p\left(\begin{pmatrix} 5 \\ x_{42} \end{pmatrix} \mid \theta \right) p(x_{42} \mid \theta_0; x_{41} = 5) dx_{42} \\
&= \int_{-\infty}^{\infty} \ln p\left(\begin{pmatrix} 5 \\ x_{42} \end{pmatrix} \mid \theta \right) \frac{p\left(\begin{pmatrix} 5 \\ x_{42} \end{pmatrix} \mid \theta^0 \right) dx_{42}}{\int_{-\infty}^{\infty} p\left(\begin{pmatrix} 5 \\ x'_{42} \end{pmatrix} \mid \theta^0 \right) dx'_{42}} \\
&= \int_{-\infty}^{\infty} \ln p\left(\begin{pmatrix} 5 \\ x_{42} \end{pmatrix} \mid \theta \right) \frac{p\left(\begin{pmatrix} 5 \\ x_{42} \end{pmatrix} \mid \theta^0 \right) dx_{42}}{\int_{0}^{10} \frac{1}{10 \times 10} dx'_{42}} \\
&= 10 \left[\int_{-\infty}^{\infty} \ln p\left(\begin{pmatrix} 5 \\ x_{42} \end{pmatrix} \mid \theta \right) p\left(\begin{pmatrix} 5 \\ x_{42} \end{pmatrix} \mid \theta^0 \right) dx_{42} \right]
\end{aligned}$$

There are 4 cases to be considered:

- $0 \le l_2 \le u_2 \le 10$ and $l_1 \le 5 \le u_1$:

$$\begin{aligned} K &= 10 \int_{l_2}^{u_2} \ln p\left(\begin{pmatrix} 5 \\ x_{42} \end{pmatrix} \mid \theta\right) \frac{1}{10 \times 10} dx_{42} \\ &= \frac{1}{10}(u_2 - l_2) \ln \frac{1}{|u_1 - l_1||u_2 - l_2|} \end{aligned}$$

- $l_2 \le 0 \le u_2 \le 10$ and $l_1 \le 5 \le u_1$:

$$\begin{aligned} K &= 10 \int_{0}^{u_2} \ln p\left(\begin{pmatrix} 5 \\ x_{42} \end{pmatrix} \mid \theta\right) \frac{1}{10 \times 10} dx_{42} \\ &= \frac{1}{10}(u_2) \ln \frac{1}{|u_1 - l_1||u_2 - l_2|} \end{aligned}$$

- $0 \le l_2 \le 10 \le u_2$ and $l_1 \le 5 \le u_1$:

$$\begin{aligned} K &= 10 \int_{l_2}^{10} \ln p\left(\begin{pmatrix} 5 \\ x_{42} \end{pmatrix} \mid \theta\right) \frac{1}{10 \times 10} dx_{42} \\ &= \frac{1}{10}(10 - l_2) \ln \frac{1}{|u_1 - l_1||u_2 - l_2|} \end{aligned}$$

- $l_2 \le u_2 \le 0 < 10$ or $0 < 10 \le l_2 < u_2$ or $5 \le l_1$ or $5 \ge u_1$ then K=0

Similarly, for L there are 4 cases:

- $0 \le l_1 < u_1 \le 10$ and $l_2 \le 6 \le u_2$:

$$\begin{aligned} L &= 10 \int_{l_1}^{u_1} \ln p\left(\begin{pmatrix} x_{51} \\ 6 \end{pmatrix} \mid \theta\right) \frac{1}{10 \times 10} dx_{51} \\ &= \frac{1}{10}(u_1 - l_1) \ln \frac{1}{|u_1 - l_1||u_2 - l_2|} \end{aligned}$$

- $l_1 \le 0 \le u_1 \le 10$ and $l_2 \le 6 \le u_2$:

$$\begin{aligned} L &= 10 \int_{0}^{u_1} \ln p\left(\begin{pmatrix} x_{51} \\ 6 \end{pmatrix} \mid \theta\right) \frac{1}{10 \times 10} dx_{51} \\ &= \frac{1}{10}(u_1) \ln \frac{1}{|u_1 - l_1||u_2 - l_2|} \end{aligned}$$

- $0 \le l_1 < 10 \le u_1$ and $l_2 \le 6 \le u_2$:

$$\begin{aligned} L &= 10 \int_{l_1}^{10} \ln p\left(\begin{pmatrix} x_{51} \\ 6 \end{pmatrix} \mid \theta\right) \frac{1}{10 \times 10} dx_{51} \\ &= \frac{1}{10}(10 - l_1) \ln \frac{1}{|u_1 - l_1||u_2 - l_2|} \end{aligned}$$

- $l_1 \le u_1 \le 0 < 10$ or $0 < 10 \le l_1 < u_1$ or $6 \le l_2$ or $6 \ge u_2$ then L=0

Therefore, $\tau(\theta, \theta^0) = \sum_{k=1}^{3} \ln p(x_k \mid \theta) + K + L$ has four different forms.

(b) [10 pts] M-Step:
The values for θ^1 are found by maximizing over the equation in part (a).

- Considering l_1:

$$\frac{\partial \sum_{k=1}^{3} \ln p(x_k \mid \theta)}{\partial l_1} = \frac{3}{u_1 - l_1} > 0 \tag{1}$$

For the three valid forms of K and L, within the defined bounds for each case, we get:

i.
$$\frac{\partial K}{\partial l_1} + \frac{\partial L}{\partial l_1} = \frac{u_2 - l_2}{10(u_1 - l_1)} + \frac{1}{10}\left[1 + \ln|u_1 - l_1||u_2 - l_2|\right] > 0 \tag{2}$$

ii.
$$\frac{\partial K}{\partial l_1} + \frac{\partial L}{\partial l_1} = \frac{u_2}{10(u_1 - l_1)} + \frac{u_1}{10(u_1 - l_1)} > 0 \tag{3}$$

iii.
$$\frac{\partial K}{\partial l_1} + \frac{\partial L}{\partial l_1} = \frac{10 - l_2}{10(u_1 - l_1)} + \frac{1}{10}\left[\frac{10 - l_1}{u_1 - l_1} + \ln|u_1 - l_1||u_2 - l_2|\right] > 0 \tag{4}$$

Therefore, combining equation (1) with either of the three cases implies that the first derivative of $\tau(\theta, \theta^0)$ with respect to l_1 is always positive. By a similar procedure, we see that the first derivative of $\tau(\theta, \theta^0)$ is also always positive with respect to l_2.

- Considering u_1:

$$\frac{\partial \sum_{k=1}^{3} \ln p(x_k \mid \theta)}{\partial u_1} = \frac{-3}{u_1 - l_1} < 0 \tag{5}$$

For the three valid forms of K and L, within the defined bounds for each case, we get:

i.
$$\frac{\partial K}{\partial u_1} + \frac{\partial L}{\partial u_1} = \frac{-(u_2 - l_2)}{10(u_1 - l_1)} + \frac{1}{10}\left[-1 - \ln(|u_1 - l_1||u_2 - l_2|)\right] < 0 \tag{6}$$

ii.
$$\frac{\partial K}{\partial u_1} + \frac{\partial L}{\partial u_1} = \frac{-u_2}{10(u_1 - l_1)} + \frac{1}{10}\left[-u_1\frac{1}{u_1 - l_1} - \ln(|u_1 - l_1||u_2 - l_2|)\right] < 0 \tag{7}$$

iii.
$$\frac{\partial K}{\partial u_1} + \frac{\partial L}{\partial u_1} = \frac{-(10 - l_2)}{10(u_1 - l_1)} + \frac{-(10 - l_1)}{10(u_1 - l_1)} < 0 \tag{8}$$

Therefore, here combining equation (5) with either of the three cases implies that the first derivative of $\tau(\theta, \theta^0)$ with respect to u_1 is always negative. By a similar procedure, we see that the first derivative of $\tau(\theta, \theta^0)$ is also always negative with respect to u_2.

Analytically, we see that the derivative of τ with respect to l_1, u_1, l_2, u_2 are always non-zero meaning that τ is monotonic in l_1, u_1, l_2, u_2 and the minimum/maximum must occur at the endpoints by verifying the second-order derivatives are positive/negative. The endpoints for l_1, u_1, l_2, u_2 are the maximum and minimum in the observed data so the M-step returns those as an estimate theta for the new round of EM.

Hence, $\theta^{(1)} = (1, 5, 2, 6)^t$

(c) [3 pts] The 2D plot should look similar to this:

(d) [3 pts] Since there is no new data available, the final θ is same as the one in part b $= (1, 5, 2, 6)^t$

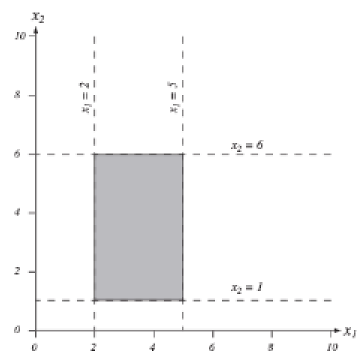

Figure 1: l1 and l2 are inter-changed. u1 and u2 limits are the same.