CSCI567 Homework 4

Question 1.1

$$P\left(x_1,\ldots,x_N\right) = P\left(x_1\right)\ldots P\left(x_N\right) = \left(\frac{1}{\theta}\right)^N$$

$$\theta = \max(x_1,\ldots,x_N)$$

θ for sure is not any x value less than the maximum of the pulled points because the probability of pulling any value greater than θ would be zero. We also have no reason to believe theta is larger than any value pulled, therefore, we set the greatest value pulled to be theta.

Question 1.2

$$P\left(k \mid x_n,\theta_1,\theta_2,\omega_1,\omega_2\right) = \frac{P\left(x_n;\theta_1,\theta_2,\omega_1,\omega_2 \mid k\right)P\left(k\right)}{P(x_n;\theta_1,\theta_2,\omega_1,\omega_2)} = \frac{P\left(x_n;\theta_1,\theta_2,\omega_1,\omega_2 \mid k\right)\omega_k}{\omega_1 U\left(X = x_n \mid \theta_1\right) + \omega_2 U\left(X = x_n \mid \theta_2\right)}$$

$$P\left(k \mid x_n,\theta_1,\theta_2,\omega_1,\omega_2\right) = \frac{\frac{\omega_k}{\theta_k}\mathbf{1}[0 < x < \theta_k]}{\omega_1 U\left(X = x_n \mid \theta_1\right) + \omega_2 U\left(X = x_n \mid \theta_2\right)}$$

$$\theta = \theta_1,\theta_2,\omega_1,\omega_2$$

$$Q\left(\theta,\theta^{old}\right) = \sum_n \sum_k p\left(k \,\middle|\, x_n;\theta^{old}\right)\log p(x_n,k|\theta)$$

$$Q\left(\theta,\theta^{old}\right) = \sum_n \sum_k \frac{\frac{\omega_k^{old}}{\theta_k^{old}}\mathbf{1}[0 < x < \theta_k^{old}]}{\omega_1^{old} U\left(X = x_n \,\middle|\, \theta_1^{old}\right) + \omega_2^{old} U\left(X = x_n \,\middle|\, \theta_2^{old}\right)}\log(\omega_k U(X = x_n|\theta_k))$$

$$A\left(\theta,\theta^{old}\right) = Q\left(\theta,\theta^{old}\right) + \sum_k H\left[p\left(k \,\middle|\, x_n;\theta^{old}\right)\right]$$

$$H\left[p\left(k \,\middle|\, x_n;\theta^{old}\right)\right] = -\int p\left(k \,\middle|\, x_n;\theta^{old}\right)\log\left(p\left(k \,\middle|\, x_n;\theta^{old}\right)\right)dx$$

$$= \int \frac{\frac{\omega_k^{old}}{\theta_k^{old}}\mathbf{1}[0 < x < \theta_k^{old}]}{\omega_1^{old} U\left(x_n \,\middle|\, \theta_1^{old}\right) + \omega_2^{old} U\left(X = x_n \,\middle|\, \theta_2^{old}\right)}\log\left(\frac{\frac{\omega_k^{old}}{\theta_k^{old}}\mathbf{1}\left[0 < x < \theta_k^{old}\right]}{\omega_1^{old} U\left(X = x_n \,\middle|\, \theta_1^{old}\right) + \text{ù}_2^{old} U\left(x_n \,\middle|\, \theta_2^{old}\right)}\right)dx$$

$$H\left[p\left(1 \,\middle|\, x_n;\theta^{old}\right)\right] = \frac{\omega_1^{old}}{\frac{\omega_1^{old}}{\theta_1^{old}} + \frac{\omega_2^{old}}{\theta_2^{old}}}\log\left(\frac{\frac{\omega_1^{old}}{\theta_1^{old}}}{\frac{\omega_1^{old}}{\theta_1^{old}} + \frac{\omega_2^{old}}{\theta_2^{old}}}\right)$$

$$H\left[p\left(1 \,\middle|\, x_n;\theta^{old}\right)\right] = \frac{\frac{\omega_2^{old}}{\theta_2^{old}}}{\frac{\omega_1^{old}}{\theta_1^{old}} + \frac{\omega_2^{old}}{\theta_2^{old}}}\log\left(\frac{\frac{\omega_2^{old}}{\theta_2^{old}}}{\frac{\omega_1^{old}}{\theta_1^{old}} + \frac{\omega_2^{old}}{\theta_2^{old}}}\right)\theta_1^{old} + \frac{\frac{\omega_2^{old}}{\theta_2^{old}}}{\frac{\omega_2^{old}}{\theta_2^{old}}}\log\left(\frac{\frac{\omega_2^{old}}{\theta_2^{old}}}{\frac{\omega_2^{old}}{\theta_2^{old}}}\right)(\theta_2^{old} - \theta_1^{old})$$

$$= \frac{\frac{\omega_2^{old}}{\theta_2^{old}}\theta_1^{old}}{\frac{\omega_1^{old}}{\theta_1^{old}} + \frac{\omega_2^{old}}{\theta_2^{old}}} \log\left(\frac{\frac{\omega_2^{old}}{\theta_2^{old}}}{\frac{\omega_1^{old}}{\theta_1^{old}} + \frac{\omega_2^{old}}{\theta_2^{old}}}\right)$$

$$\theta^{new} = \max_{\theta} A\left(\theta, \theta^{old}\right)$$

$$\theta_1^{new} = \max_{\theta_1} \sum_n P_{old}(k = 1|x_n)\log(\omega_1 U(X = x_n|\theta_1))$$

$$\theta_2^{new} = \max_{\theta_2} \sum_n P_{old}(k = 2 \mid x_n)\log(\omega_2 U(X = x_n \mid \theta_2))$$

Since the rest of the variables do not depend on θ_1 or θ_2.

Question 2.1

$$P(x_b \mid x_a) = \frac{P(x_a,\ x_b)}{P(x_a)}$$

$$= \frac{P(x)}{P(x_a)}$$

$$= \frac{\sum_{k=1}^{K} \pi_k P(x_a, x_b \mid k)}{\sum_{j=1}^{K} \pi_j P(x_a|j)}$$

$$= \frac{\sum_{k=1}^{K} \pi_k P(x_b \mid x_a k) P(x_a|k)}{\sum_{j=1}^{K} \pi_j P(x_a|j)}$$

$$P(x_b \mid x_a) = \sum_{k=1}^{K} \frac{\pi_k P(x_a \mid k)}{\sum_{j=1}^{K} \pi_j P(x_a|j)} P(x_b \mid x_a k)$$

$$\lambda_k = \frac{\pi_k P(x_a \mid k)}{\sum_{j=1}^{K} \pi_j P(x_a|j)}$$

Question 3.1

$$\lim_{\sigma \to 0}(\gamma(z_{nk})) = \lim_{\sigma \to 0} \frac{\pi_k \exp\left(-\frac{\|x_n - \mu_k\|^2}{2\sigma^2}\right)}{\sum_j \pi_j \exp\left(-\frac{\|x_n - \mu_j\|^2}{2\sigma^2}\right)}$$

As σ goes to zero, the Gaussian becomes a dirac-delta function.

$$= \frac{\pi_k \delta(x_n - \mu_k)}{\sum_j \pi_k \delta\left(x_n - \mu_j\right)} = \begin{cases} \frac{\pi_k \delta(0)}{\pi_k \delta(0)} = 1, & if\ x_n = \mu_k \\ 0, & if\ x_n \neq \mu_k \end{cases}$$

$$\lim_{\sigma \to 0}(\gamma(z_{nk})) = r_{nk}$$

$$\lim_{\sigma \to 0} \sum_n^N \sum_k^K \gamma(z_{nk})[\log \pi_k + \log N(x_n|\mu_k, \sigma^2 I)] = \sum_n^N \sum_k^K r_{nk}[\log \pi_k + \log \delta(x_n - \mu_k)]$$

Maximizing this function with the delta function, is the same as minimizing the distances between x_n and μ_k.

$$\max_{\{\mu_k\}} \sum_{n}^{N} \sum_{k}^{K} r_{nk}[\log \pi_\mathrm{k} + \log \delta(\mathrm{x_n} - \mu_\mathrm{k})] = \min_{\{\mu_k\}} \sum_{n}^{N} \sum_{k}^{K} r_{nk} \left\| \mathrm{x_n} - \mu_\mathrm{k} \right\|_2^2$$

Question 4.1

$$\frac{1}{\sqrt{2\pi\sigma_{y_n n}^2}} \exp(-\frac{\left(x_d - \mu_{y_n n}\right)^2}{2\sigma_{y_n n}^2})$$

$$\log\left(P\left(D\right)\right) = \log\left(\prod_{n}^{N} \pi_{y_n} \prod_{k}^{C} \left(\frac{1}{\sqrt{2\pi\sigma_{y_n k}^2}} exp\left(-\frac{\left(x_d - \mu_{y_n k}\right)^2}{2\sigma_{y_n k}^2}\right)\right)^{z_{nk}}\right)$$

Where z gives the number of times k in x

$$= \sum_{n}^{N} \log\left(\pi_{y_n}\right) + \sum_{n,k}^{N,C} z_{nk} \log \frac{1}{\sqrt{2\pi\sigma_{y_n k}^2}} exp\left(-\frac{\left(x_d - \mu_{y_n k}\right)^2}{2\sigma_{y_n k}^2}\right)$$

$$= \sum_{n}^{N} \log\left(\pi_{y_n}\right) + \sum_{k}^{C} z_{nk} \log \frac{1}{\sqrt{2\pi\sigma_{y_n k}^2}} + \log exp\left(-\frac{\left(x_d - \mu_{y_n k}\right)^2}{2\sigma_{y_n k}^2}\right)$$

$$log - likelihood = \sum_{n}^{N} \log\left(\pi_{y_n}\right) + \sum_{k}^{C} z_{nk} \log \frac{1}{\sqrt{2\pi\sigma_{y_n k}^2}} - \frac{\left(x_d - \mu_{y_n k}\right)^2}{2\sigma_{y_n k}^2}$$

Question 4.2

To estimate each individual Gaussian, there is no reason to believe that the mean of the Gaussian is different from the mean of the points labeled as being in the Gaussian.

$$\mu_{ck} = \frac{\left(\sum_{n:\ y_n = c} x_n\right)}{\#points\ labeled\ c}$$

Similarly, standard deviation should also be the standard deviation of the points within the same label:

$$\sigma_{ck} = \sqrt{\frac{\sum_{n:\ y_n = c} \left(x_n - \mu_{ck}\right)^2}{\left(\#points\ labeled\ c\right) - 1}}$$

The pi should be no different from the spam/ham email case:

$$\pi_c = \frac{\#points\ labeled\ as\ c}{N}$$