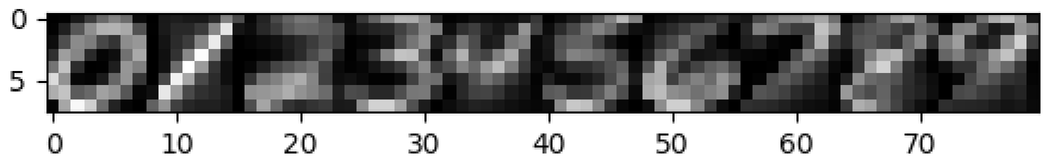
Q1.

a) Average Conditional Likelihood for training set is: -0.124624436669
Average Conditional Likelihood for test set is: -0.196673203255

b) Accuracy for training set is: 0.9814285714285714
Accuracy for test set is: 0.97275

c)




Q2.

a)

$$P(\theta) \propto \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1}$$

$$p(X_i = j \mid \theta) = \theta_j$$

$$P(D \mid \theta) = \prod_{i=1}^{n} p(x_i = x_i \mid \theta) = \prod_{i=1}^{n} \theta_{x_i} = \prod_{i=1}^{n} \prod_{j=1}^{m} \theta_j^{I(x_i = j)}$$

$$= \prod_{j=1}^{n} \theta_j^{\sum I(x_i = j)}$$

let $N = \sum I(x_i = j)$

$$= \prod_{j=1}^{n} \theta_j^{N}$$

$$P(\theta \mid D) \propto P(\theta) P(D \mid \theta)$$

$$= \prod_{K=1}^{K} \theta_K^{\alpha_K - 1} \prod_{K=1}^{K} \theta_K^{N}$$

$$= \prod_{K=1}^{K} \theta_K^{\alpha_K - 1 + N}$$

$$P(D' \mid D) = \theta_{pred} = \int p(\theta \mid D)\, p(D' \mid D)\, d\theta$$

$$= \int \text{Dirchlet}(\alpha_1 + N_1, \ldots, \alpha_k + N_k)\, \theta_K \, d\theta$$

$$= E(\theta_k \mid D)$$

$$= \frac{N_k + \alpha_k}{N + \sum \alpha_k}$$

b) $\hat{\theta}_{MAP} = \text{argmax } P(\theta | D)$

$= \text{argmax } P(D|\theta) P(\theta)$

$= \text{argmax } \log P(\theta) + \log P(D|\theta)$

# $\log P(D|\theta) = \log \prod_{k=1}^{K} \theta_k^{\alpha_k + N - 1}$

$= \sum_{k=1}^{K} (\alpha_k + N - 1) \log \theta_k$

# $\frac{\partial f}{\partial \theta_i} = \frac{\alpha_i + N_i - 1}{\theta_i}$ * $\frac{\partial g}{\partial \theta_i} = 1$

$= \lambda \frac{\partial g}{\partial \theta_i}$

$\frac{\alpha_i + N_i - 1}{\theta_i} = \lambda$

$\Sigma(\alpha_i + N_i - 1) = \lambda \Sigma \theta_i$

$MAP = \frac{\alpha + N - 1}{N - K + \sum_{j=1}^{K} \alpha_j}$

Q3.

a) E - step:

$q(z=k) = P(z=k | x, \theta) = \frac{P(x|z, \theta) \, P(z|\theta)}{\Sigma_{j=0}^{1} P(x|z_j, \theta) \, P(z_j | \theta)}$

$= \frac{[\prod_i P(x_i|z, \theta)] P(z|\theta)}{\Sigma_{j=0}^{1} P(x|z_j, \theta) \, P(z_j|\theta)}$

$= \frac{[\prod_i P(x_i|\mu, \sigma)] \, (\pi^k (1-\pi)^{(1-k)})}{\Sigma_{j=0}^{1} \prod_i P(x_i|\mu, \sigma) \, \pi^j (1-\pi)^{(1-j)}}$

$\mu(z) = E(z|x, \theta) = \frac{\pi \prod_{d=1}^{D} (\theta_d)^{x_d} (1-\theta_d)^{1-x_d}}{\Sigma_j \pi_j \prod_{d=1}^{D} (\theta_d^j)^{x_d} (1-\theta_d^j)^{1-x_d}}$

b) set derivation to 0:

$$\frac{d}{d\mu} E[\log P(z, D, \mu)] = \sum_{i=1}^{N} E(z|x,\theta)\left[\frac{x}{\mu} + \frac{1-x}{1-\mu}\right] = 0$$

$$\sum_{i=1}^{N} E(z|x,\theta)\left[x(1-\mu) + (1-x)\mu\right] = \sum_{i=1}^{N} E(z|x,\theta)\left[-\mu + x\right] = 0$$

$$\mu = \frac{\sum_{i=1}^{N} q(z^{(i)})\, x^{(i)}}{\sum_{i=1}^{N} q(z^{(i)})}$$

$$\mu = \frac{\sum_{i=1}^{N} m^{(i)}\, x^{(i)}}{N}$$