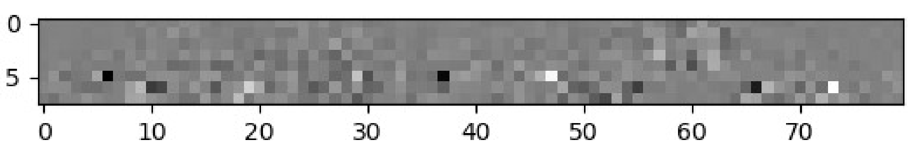
Q1 a)
train set -- average conditional log-likelihood is: -0.12458797208729222
test set -- average conditional log-likelihood is: -0.19660908967370488

b)
train set -- accuracy is: 0.9814285714285714
test set -- accuracy is: 0.97275

c)



Q2 a) $P(\theta) \propto \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1}$

$= \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$

$P(X_i = j | \theta) = \theta_j$

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

let $N = \Sigma I(X_i = j)$ $\quad = \prod_{j=1}^{n} \theta_j^{\Sigma I(X_i = j)}$

$P(\theta|D) \propto P(\theta) P(D|\theta)$ $\quad = \prod_{j=1}^{n} \theta_j^{N}$

$= \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} \prod^{K} \theta_k^{N}$

$= \prod_{k=1}^{K} \theta_k^{\alpha_k + N - 1}$

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

$= \int \text{Dirchlet}(\alpha_1 + N_1, \cdots, \alpha_k + N_k) \; \theta_k \, d\theta.$

$= E(\theta_k | D.$

$= \dfrac{N_k + \alpha_k}{N + \Sigma \alpha_k}.$

b) $\hat{\theta}_{MAP} = \arg\max P(\theta|D)$

$= \arg\max P(D|\theta) P(\theta).$

$= \arg\max \log P(\theta) + \log P(D|\theta)$

$\log P(\theta|D)$

$= \log \prod_{k=1}^{K} \theta_k^{\alpha_k + N - 1}$

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

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

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

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

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

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