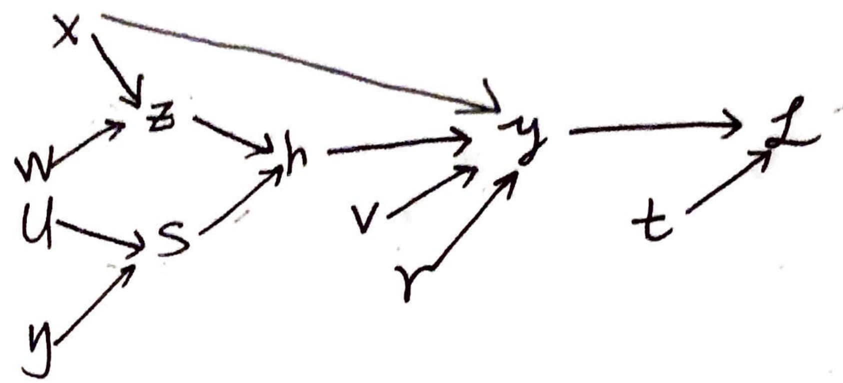
1.

a)



b) $\bar{y} = y - t$, $\bar{t} = y - t$

$\bar{h} = \bar{y}\, v$, $\bar{v} = \bar{y}\, h$, $\bar{r} = \bar{y}\, x$

$\bar{z} = \bar{h}\,(\sigma(s) \odot)$, $\bar{s} = \bar{h}\, \sigma'(s)\,(z \odot)$

$\bar{u} = \bar{s}\, \eta^{T}$, $\bar{\eta} = \bar{s}\, u^{T}$, $\bar{w} = \bar{z}\, x^{T}$

$\bar{x} = \bar{z}\, W^{T} + \bar{y}\, r$

So $\bar{W} = (y-t)\, v\, (\sigma(s) \odot)(x^{T})$

$\bar{u} = (y-t)\, v\, \sigma'(s)\, (z \odot)(\eta^{T})$

$\bar{v} = (y-t)\, h$, $\bar{r} = (y-t)\, x$

$\bar{x} = (y-t)\, v\, (\sigma(s) \odot)(W^{T}) + (y-t)\, r$

$\bar{\eta} = (y-t)\, v\, \sigma'(s)\, (z \odot)\, u^{T}$

2.a) $\ell(\theta) = \sum_{i=1}^{N} \log(p(x^{(i)}, c^{(i)} \mid \theta, \pi))$

$$= \sum_{i=1}^{N} \left(\log(p(c^{(i)} \mid \pi)) + \sum_{j=1}^{784} \log(p(x_j^{(i)} \mid c^{(i)}, \theta_{jc^{(i)}})) \right)$$

$$= \sum_{i=1}^{N} \left(\log(\pi_{c^{(i)}}) + \sum_{j=1}^{781} \left(x_j^{(i)} \log(\theta_{jc^{(i)}}) + (1 - x_j^{(i)}) \log(1 - \theta_{jc^{(i)}}) \right) \right)$$

$$\frac{\partial \ell}{\partial \theta_{jc}} = \sum_{i=1}^{N} \mathbb{I}(c^{(i)} = c) \left(\frac{x_j^{(i)}}{\theta_{jc}} - \frac{1 - x_j^{(i)}}{1 - \theta_{jc}} \right) = 0$$

$$\Rightarrow \sum_{i=1}^{N} \mathbb{I}(c^{(i)} = c) \left(x_j^{(i)} - \theta_{jc} x_j^{(i)} - \theta_{jc} + \theta_{jc} x_j^{i} \right) = 0$$

$$\cancel{\Rightarrow \hat{\theta}_{jc} = \frac{\sum_{i=1}^{N} \mathbb{I}(c^{(i)} = c \,\&\, x_j^{(i)} = 1)}{\sum_{i=1}^{N} \mathbb{I}(c^{(i)} = c)}}$$

$$\Rightarrow \sum_{i=1}^{N} \mathbb{I}(c^{i} = c)\, x_j^{(i)} = \left(\sum_{i=1}^{N} \mathbb{I}(c^{(i)} = c) \right) \theta_{jc}$$

$$\Rightarrow \hat{\theta}_{jc} = \frac{\sum_{i=1}^{N} \mathbb{I}(c^{(i)} = c)\, x_j^{(i)}}{\sum_{i=1}^{N} \mathbb{I}(c^{(i)} = c)}$$

$$\ell(\pi) = \sum_{i=1}^{N} \log \prod_{j=0}^{9} \pi_j^{t_j^{(i)}} = \sum_{i=1}^{N} \sum_{j=0}^{9} t_j^{(i)} \log \pi_j$$

$$= \sum_{i=1}^{N} \left(\sum_{j=1}^{8} t_j^{(i)} \log \pi_j + t_9^{(i)} \log(\pi_9) \right)$$

$$\frac{\partial}{\partial \pi_j}(\log \pi_9) = -\frac{1}{\pi_9}$$

$$\frac{\partial \ell}{\partial \pi_j} = \sum_{i=1}^{N} \frac{t_j^{(i)}}{\pi_j} - \frac{t_9^{(i)}}{\pi_9} = 0$$

$$\Rightarrow \sum_{i=1}^{N} \frac{t_j(i)}{\pi_j} = \sum_{i=1}^{N} \frac{t_9^{(i)}}{\pi_9}$$

$$\Rightarrow \hat{\pi}_j = \frac{\sum_{i=1}^{N} t_j(i)}{\ldots}$$

Since $\sum \pi_j = 1$,

$$\frac{1}{\hat{\pi}_9} = \sum_{j=1}^{9} \frac{\hat{\pi}_j}{\hat{\pi}_9} = \sum_{j=1}^{9} \frac{\sum_{i=1}^{N} t_j^{(i)}}{\sum_{i=1}^{N} t_9^{(i)}} = \frac{N}{\sum_{i=1}^{N} t_9^{(i)}}$$

$$\Rightarrow \quad \hat{\pi}_9 = \frac{\sum_{i=1}^{N} t_9^{(i)}}{N}$$

$$\hat{\pi}_j = \frac{\sum_{i=1}^{N} t_j^{(i)}}{N}$$

b) $$\log p(t \mid x, \theta, \pi) = \log\left(\frac{p(t \mid \theta, \pi)\, p(x \mid t, \theta, \pi)}{p(x \mid \theta, \pi)}\right)$$

$$= \log\left(\frac{p(c \mid \pi) \prod_{j=1}^{784} p(x_j \mid c, \theta, \pi)}{\sum_{k=0}^{9} p(c_k \mid \theta, \pi)\, p(x \mid c_k, \theta, \pi)}\right)$$

$$= \log\left(\frac{\pi_c \prod_{j=1}^{784} \theta_{jc}^{x_j} (1-\theta_{jc})^{1-x_j}}{\sum_{k=0}^{9} \pi_k \prod_{i=1}^{784} \theta_{jk}^{x_i} (1-\theta_{jk})^{1-x_i}}\right)$$

$$= \log \pi_c + \sum_{j=1}^{784} x_j \log \theta_{jc} + (1-x_j) \log(1-\theta_{jc})$$
$$- \log\left(\sum_{k=0}^{9} \pi_k \prod_{i=1}^{784} \theta_{jk}^{x_i} (1-\theta_{jk})^{1-x_i}\right)$$

$$= \log \pi_c + \sum_{j=1}^{784} x_j \log \theta_{jc} + (1-x_j) \log(1-\theta_{jc})$$
$$- \log\left(\sum_{k=0}^{9} \pi_k \exp\left(\sum_{i=1}^{784} x_i \log(\theta_{jk}) + (1-x_i) \log(1-\theta_{jk})\right)\right)$$

c) The average log-likelihood is nan because $\log(\theta_{jk})$ is undefined for $\theta_{jk}=0$.

d) First image

e) $\theta_{jc} \sim \text{Beta}(3,3)$, $p(\theta|x,t,\pi)\cdot p(x,t,\pi) = p(\theta,x,t,\pi)$

$$= p(\theta)p(x,t,\pi|\theta)$$

So $p(\theta|x,t,\pi) = C\cdot p(\theta)p(x,t|\theta,\pi) = p(x,t|\theta,\pi)\,p(\theta,\pi)$

$$l(\theta) = \sum_{i=1}^{N} \log p(\theta^{(i)}) + \log p(x^{(j)}, t^{(i)}|\theta^{(i)},\pi) + C$$

$$= \sum_{i=1}^{N} \log \theta_{jc}^{(i)} + \log(1-\theta_{jc}^{(i)}) + \log \pi_c^{(i)} + \sum_{j=1}^{784} x_j^{(i)} \log \theta_{jc}^{(i)} + (1-x_j^{(i)}) \log (1-\theta_{jc}^{(i)}) + C$$

So $\frac{\partial l}{\partial \theta_{jc}} = \frac{1}{\theta_{jc}} - \frac{1}{1-\theta_{jc}} + \sum_{i=1}^{N} \mathbb{I}(c_i=c)\left(\frac{x_j^{(i)}}{\theta_{jc}} - \frac{1-x_j^{(i)}}{1-\theta_{jc}}\right)$

$$= 0$$

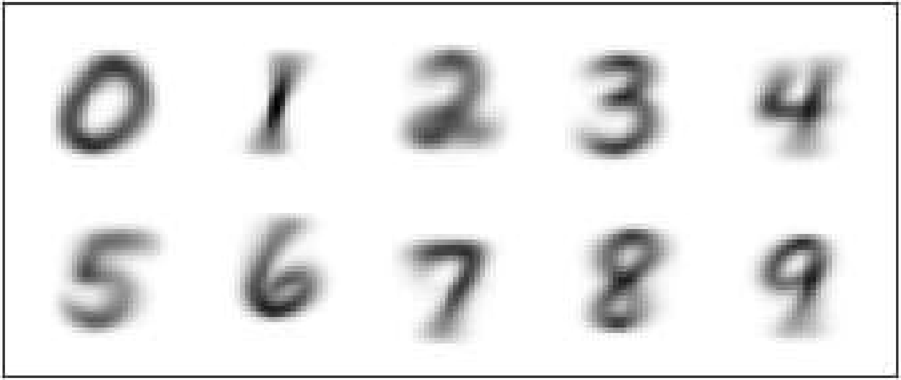
$$\Rightarrow \hat{\theta}_{jc} = \frac{\sum_{i=1}^{N} \mathbb{I}(c^{(i)}=c)\, x_j^{(i)} + 2}{\sum_{i=1}^{N} \mathbb{I}(c^{(i)}=c) + 4}$$

f) average log-likelihood : -3.3571

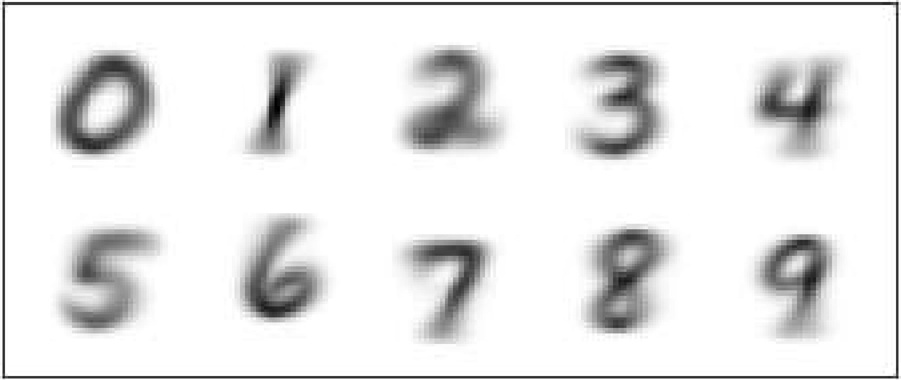
training accuracy : 0.8352

test accuracy : 0.816

g) Second image.



MLE



MAP

3a)

$$p(\theta) \propto \theta_1^{a_1-1} \cdots \theta_K^{a_K-1}$$

$$p(\mathcal{D}|\theta) = \prod_{i=1}^{N} p(x^{(i)}|\theta)$$

$$= \prod_{i=1}^{N} \left(\prod_{k=1}^{K} \theta_k^{x_k^{(i)}} \right)$$

$$= \prod_{k=1}^{K} \theta_k^{N_k}$$

So $p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)\, p(\theta)$

$$\cancel{= \prod_{i=1}^{N} \left(\prod_{k=1}^{K} \theta_k^{x_k^{(i)}} \right) \cdot \prod}$$

$$= \prod_{i=1}^{K} \theta_k^{N_k} \cdot \prod_{j=1}^{K} \theta_j^{a_j-1}$$

$$= \prod_{i=1}^{K} \theta_k^{N_k+a_k-1}$$

So yes, the Dirichlet distribution is a conjugate prior.

b) $l(\theta) = \log(p(\theta|\mathcal{D}))$

$$= \log\left(\prod_{i=1}^{K} \theta_k^{N_k+a_k-1} \right) + c$$

$$= \sum_{i=1}^{K} (N_k+a_k-1) \log \theta_k + c$$

$$\cancel{\frac{\partial l}{\partial \theta_k} = \frac{N_k+a_k-1}{\theta_k}}$$

$$= \sum_{i=1}^{K-1} (N_k+a_k-1) \log \theta_k + (N_K+a_K-1) \log \theta_K$$

$$\frac{\partial l}{\partial \theta_k} = \frac{N_k+a_k-1}{\theta_k} - \frac{N_K+a_K-1}{\theta_K} = 0$$

$$\Rightarrow \frac{\hat{\theta}_K}{\hat{\theta}_k} = \frac{N_K+a_K-1}{N_k+a_k-1}, \quad \frac{\hat{\theta}_k}{\hat{\theta}_K} = \frac{N_k+a_k-1}{N_K+a_K-1}$$

$$\Rightarrow \frac{1}{\hat{\theta_k}} = \sum_{i=1}^{K} \frac{\hat{\theta_k}}{\hat{\theta_K}} = \frac{\sum_{k=1}^{K} N_k + a_k - 1}{N_k + a_k - 1}$$

$$= \frac{N - K + \sum_{k=1}^{K} a_k}{N_k + a_k - 1}$$

$$\Rightarrow \hat{\theta_K} = \frac{N_K + a_K - 1}{N - K + \sum_{k=1}^{K} a_k}$$

c) Let $\theta^{(p)} = \theta \mid \mathcal{D}$, i.e. $p(\theta^{(p)}) \overset{= p(\theta \mid \mathcal{D})}{\sim}$ Dirichlet$(N_1 + a_1,$ $\cdots N_K + a_K)$

$$p(x_k^{(N+1)} = 1 \mid \theta^{(p)}) = \mathbb{E}(\theta_k^{(p)})$$

$$= \frac{N_k + a_k}{\sum_{k'} N_{k'} + a_{k'}}$$

$$= \frac{N_k + a_k}{N + \sum_{k'} a_{k'}}$$

4a)

$$p(y|x,\mu,\Sigma)\, p(x|\mu,\Sigma) = p(x|y,\mu,\Sigma)\, \overset{\swarrow \frac{1}{10}}{p(y|\mu,\Sigma)}$$

$$\Rightarrow \quad p(y|x,\mu,\Sigma) = \frac{1}{10} \frac{p(x|y,\mu,\Sigma)}{\sum_{j=1}^{10} p(x|y^{(j)},\mu,\Sigma)}$$

So $\cancel{\log}$ $\frac{1}{10}\sum_{i=1}^{10} \log(p(y^{(i)}|x^{(i)},\mu,\Sigma)$

$$= \frac{1}{10}\sum_{i=1}^{10} \log\left(\frac{1}{10}\right) + \log(p(x^{(i)}|y^{(i)},\mu,\Sigma)) - \sum_{j=1}^{10} p(x^{(i)}|y^{(j)},\mu,\Sigma)$$

$p(x|y,\mu,\Sigma)$ is given by the formula.

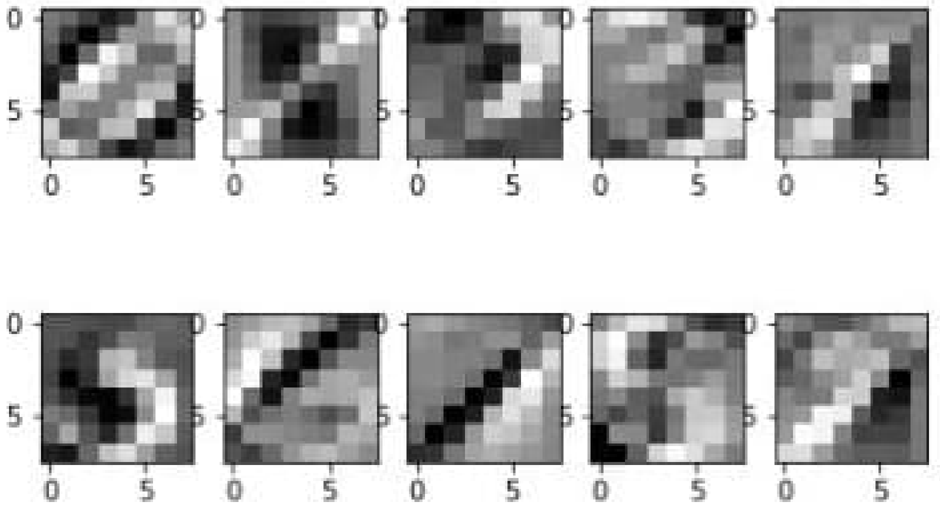
The arg ~~like~~ Conditional log-likelihood for training set is $-0.45150334\ldots$

for test set is $-1.574084784\ldots$

b) Training ~~a~~ accuracy is $0.9862857\ldots$

test accuracy is 0.95925

c) The graph is attached.



0
5
0
5