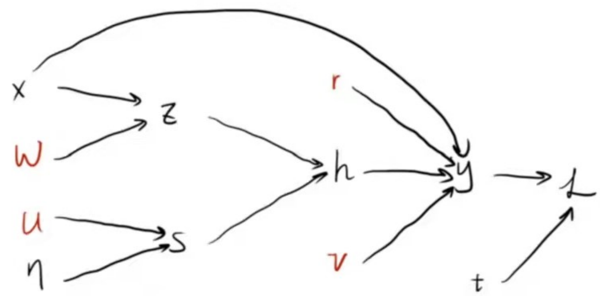
Q1



a)

b)

$$\overline{y} = y - t$$

$$\overline{\mathbf{v}} = \overline{y}\mathbf{h}^T$$

$$\overline{\mathbf{r}} = \overline{y}\mathbf{x}^T$$

$$\overline{\mathbf{h}} = \overline{y}\mathbf{v}^T$$

$$\overline{\mathbf{z}} = \overline{\mathbf{h}}\sigma(\mathbf{s})$$

$$\overline{\mathbf{s}} = \overline{\mathbf{h}}(z \odot \sigma'(\mathbf{s}))$$

$$\overline{\mathbf{W}} = \overline{\mathbf{z}}\mathbf{x}$$

$$\overline{\mathbf{U}} = \overline{\mathbf{s}}\eta$$

$$\overline{\eta} = \overline{\mathbf{s}}\mathbf{U}$$

$$\overline{\mathbf{x}} = \overline{\mathbf{z}}\mathbf{W} + \overline{\mathbf{y}}\mathbf{r}^T$$

Q2

a) For each x, write

$$p(x,t|\theta,\pi) = \prod_{i=0}^{9}[\pi_i^{t_i}\prod_{j=1}^{784}(\theta_{ji})^{x_j}(1-\theta_{ji})^{(1-x_j)}]$$

Assume $\{x_1,\cdots,x_N\}$ are all the N samples, let x_{Nj}, t_{Nj} be the x_j, t_j value of the Nth sample, the likelihood function is:

$$L(\theta,\pi) = \prod_{k=1}^{N}\prod_{i=0}^{9}[\pi_i^{t_{ki}}\prod_{j=1}^{784}(\theta_{ji})^{x_{kj}}(1-\theta_{ji})^{(1-x_{kj})}]$$

Take logarithm:

$$l(\theta,\pi) = \sum_{k=1}^{N}\sum_{i=0}^{9}[t_{ki}log(\pi_i) + \sum_{j=1}^{784}x_{kj}log(\theta_{ji}) + (1-x_{kj})log(1-\theta_{ji})]$$

To find MLE of θ, take derivative w.r.t θ_{ji} and set to 0, we get

$$\sum_{k=1}^{N}t_{ki}(\frac{x_{kj}}{\theta_{ji}} - \frac{1-x_{kj}}{1-\theta_{ji}}) = 0$$

Multiply both side by $\theta(1-\theta)$, simplify we get

$$\hat{\theta}_{ji} = \frac{\sum_{k=1}^{N}t_{ki}x_{kj}}{\sum_{k=1}^{N}t_{ki}}$$

To find MLE of π, first write $l(\theta,\pi) = \sum_{k=1}^{N}(\sum_{i=0}^{8}t_{ki}log(\pi_i) + t_{k9}log(1-\sum_{i=0}^{8}\pi_i) + \cdots)$, take derivative w.r.t π_i and set to 0:

$$\sum_{k=1}^{N}(\frac{t_{ki}}{\pi_i} - \frac{t_{k9}}{1-\sum_{i=0}^{8}\pi_i}) = 0$$

Multiply both sides by $\pi_i(1-\pi_9)$, simplify we get

$$\frac{\pi_i}{\pi_9} = \frac{\sum_{k=1}^{N} t_{ki}}{\sum_{k=1}^{N} t_{k9}}$$

Sum up the functions for 10 values of is, note sum of all π_i is 1, and sum of t_{ki} over all k and i is N, so we get

$$\frac{1}{\pi_9} = \frac{N}{\sum_{k=1}^{N} t_{k9}}$$

Do this manipulation over all values of i, we get MLE of π_i

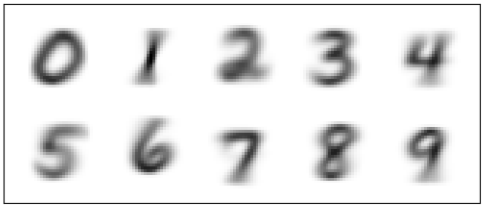
$$\hat{\pi}_i = \frac{\sum_{k=1}^{N} t_{ki}}{N}$$

b) We can write

$$\begin{aligned}
\log(p(t|x,\theta,\pi)) &= \log(\frac{p(x,t|\theta,\pi)}{p(x|\theta,\pi)}) \\
&= \log(\frac{p(t|\pi)\prod_{j=1}^{784}(x_j|t,\theta)}{\sum_{i=0}^{9} p(t_i|\pi)\prod_{j=1}^{784}(x_j|t,\theta)}) \\
&= \log(\pi_c) + \sum_{j=1}^{784}(x_j \log(\theta_{jc}) + (1-x_j)\log(1-\theta_{jc})) \\
&- \log(\sum_{i=0}^{9} \pi_i \prod_{j=1}^{784} \theta_{ji}^{x_j}(1-\theta_{ji})^{1-x_j})
\end{aligned}$$

c) The average log-likelihood is nan, because at some point the log is taken over 0 maximum likelihood, which results in infinity error.

d) MLE estimator plot:



e) MAP estimator:

$$\log p(\theta, D) = \log p(\theta) + \log p(D|\theta)$$

$$\log P(\theta_{jc}|B(3,3)) = \log\left(\frac{\Gamma(6)}{\Gamma(3)\Gamma(3)}\,\theta_{jc}^{2}\,(1-\theta_{jc})^{2}\right) = \log\left(\frac{\Gamma(6)}{\Gamma(3)\Gamma(3)}\right) + 2\log\theta_{jc} + 2\log(1-\theta_{jc})$$

and we know $\mathcal{L} = \log P(D|\theta)$ and $\frac{d\mathcal{L}}{\partial\theta}$

$$\frac{d}{d\theta_{jc}} \log p(\theta, D) = \frac{2}{\theta_{jc}} - \frac{2}{1-\theta_{jc}} + \sum_{k=1}^{N} t_{kc}\left(\frac{x_{kj}}{\theta_{jc}} - \frac{1-x_{kj}}{1-\theta_{jc}}\right) = 0$$

$$2(1-\theta_{jc}) - 2\theta_{jc} + \sum_{k=1}^{N} t_{kc}\left(x_{kj}(1-\theta_{jc}) - (1-x_{kj})\theta_{jc}\right) = 0$$

$$2 - 4\theta_{jc} + \sum_{k=1}^{N} t_{kc}(x_{kj} - \theta_{jc}) = 0$$

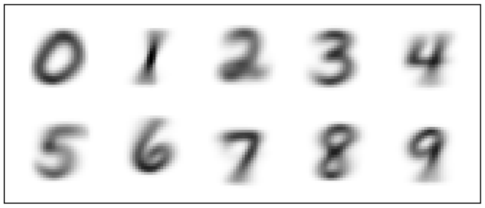
$$\hat{\theta}_{jc} = \frac{2 + \sum_{k=1}^{N} t_{kc} x_{kj}}{4 + \sum_{k=1}^{N} t_{kc}}$$

f) Average log-likelihood for MAP is -3.357

Training accuracy for MAP is 0.835

Test accuracy for MAP is 0.816

g) MAP estimator plot:



Q3

a)

$$p(\theta|D) \propto p(\theta)p(D|\theta)$$

$$\propto \left(\theta_1^{a_1-1} \cdots a_K^{a_K-1}\right) \cdot \prod_{i=1}^{N} \prod_{k=1}^{K} \theta_k^{x_k}$$

$$= (\theta_1^{a_1-1} \cdots a_K^{a_K-1}) \prod_{k=1}^{K} \theta_k^{N_k} \qquad N_k = \sum_{i=1}^{N} x_k^{(i)}$$

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

b)

$$l(\theta) = \sum_{k=1}^{K} (a_k + N_k - 1)\log(\theta_k)$$

$$= \sum_{i=1}^{K-1} (a_i + N_i - 1)\log(\theta_i) + (a_K + N_K - 1)\log(1 - \sum_{j=1}^{K-1}\theta_j)$$

$$\frac{dl}{d\theta_i} = \frac{a_i + N_i - 1}{\theta_i} - \frac{a_K + N_K - 1}{\theta_K} = 0$$

$$\frac{\theta_K}{\theta_i} = \frac{a_K + N_K - 1}{a_i + N_i - 1} \qquad \text{sum over } \theta_i :$$

$$\hat{\theta}_K = \frac{a_K + N_K - 1}{\sum_{i=1}^{K}(a_i + N_i - 1)}$$

Do this for θ_i, so we have

$$\hat{\theta}_i = \frac{a_i + N_i - 1}{\sum_{i=1}^{K}(a_i + N_i - 1)}$$

c)

$$p(x^{N+1} = 1 | D) = \int p(x^{N+1} = k | \theta) p(\theta | D) d\theta$$

$$= \int \theta_k \, p(\theta|D) \, d\theta$$

$$= E(\theta_k | D) = \frac{a_k + N_k}{\sum_{i=1}^{K} a_i + N_i}$$