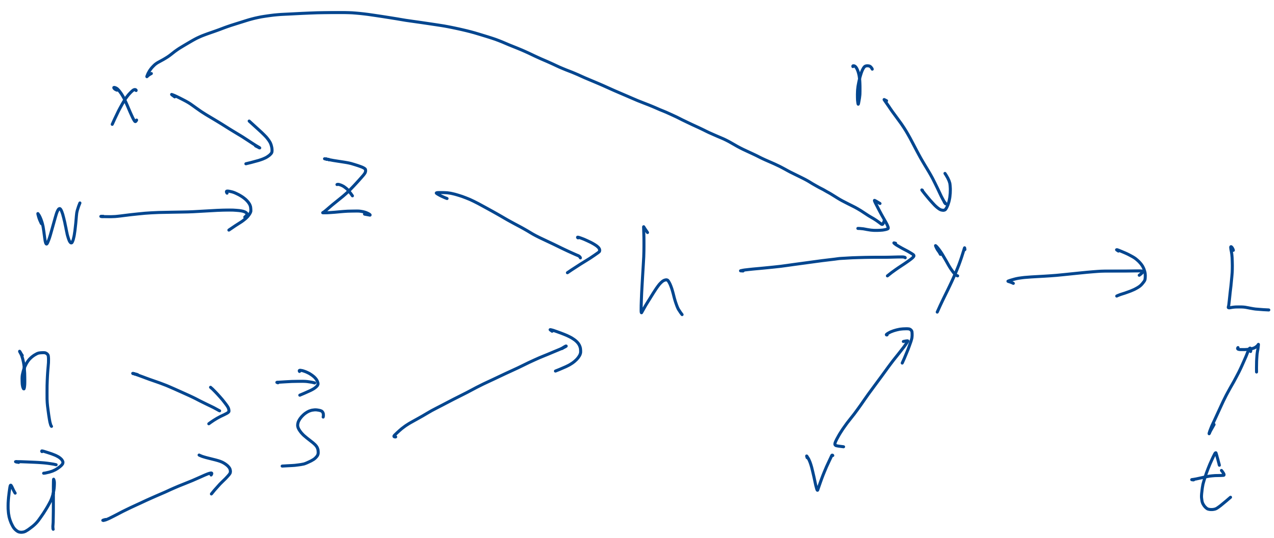
1.

a



b

$$\bar{L} = 1$$

$$\bar{t} = \bar{L} \cdot \frac{dL}{dt}$$

$$\bar{y} = \bar{L} \, \frac{dL}{dy} = y - t$$

$$\bar{t} = \bar{L} \, \frac{dL}{dt} = t - y$$

$$\bar{r} = \bar{y} \cdot \frac{dy}{dr} = (y - t) \cdot x$$

$$\bar{v} = \bar{y} \cdot \frac{dy}{dv} = (y - t) \cdot h$$

$$\bar{h} = \bar{y} \cdot \frac{dy}{dh} = (y - t) \cdot v$$

$$\bar{z} = \bar{h} \cdot \frac{dh}{dz} = (y - t) \cdot v \cdot \sigma(s)$$

$$\bar{s} = \bar{h} \, \frac{dh}{ds} = (y - t) \cdot v \cdot z \cdot \sigma'(s)$$

$$\bar{u} = \bar{s} \cdot \frac{ds}{du} = (y - t) v z \sigma'(s) \cdot \eta$$

$$\bar{\eta} = \bar{s} \cdot \frac{ds}{d\eta} = (y - t) v z \sigma'(s) \cdot u$$

$$\bar{w} = \bar{z} \cdot \frac{dz}{dw} = (y - t) v \cdot \sigma(s) \cdot x$$

$$\bar{x} = \bar{z} \frac{dz}{dx} + \bar{y} \frac{dy}{dx} = (y - t) \cdot v \cdot \sigma(s) \cdot w + (y - t) \cdot r$$

2.

a

MLE by θ, π can be expressed:

$$L(\theta,\pi) = p(c|\pi) \prod_{j=1}^{784} p(x_j | c, \theta_{jc})$$

$$l(\theta,\pi) = \underbrace{\log(p(c|\pi))}_{①} + \underbrace{\sum_{j=1}^{784} \log(p(x_j|c,\theta_{jc}))}_{②}$$

Do partial derivatives on ①, ② separatly, π in ①, θ in ②

$$①: \log(p(c|\pi)) = \sum_{i=1}^{N} \sum_{j=0}^{9} \log(\pi_j^{t_j^{(i)}})$$

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

$$= \sum_{i=1}^{N} (1-\sum_{j=0}^{8} \pi_j) t_9^{(i)} + \sum_{j=0}^{8} t_j^{(i)} \log \pi_j$$

$\forall j < 9$

$$\frac{\partial (1)}{\partial \pi_j} = \sum_{i=1}^{N} \frac{-t_9^{(i)}}{1-\sum_{j=0}^{8} \pi_j} + \frac{t_j^{(i)}}{\pi_j} = 0$$

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

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

So $$\frac{\hat{\pi}_j}{\hat{\pi}_9} = \frac{\sum_{i=1}^{N} \mathbb{I}(t_j^{(i)}=1)}{\sum_{i=1}^{N} \mathbb{I}(t_9^{(i)}=1)}$$

by $\sum_{j=0}^{9} \pi_j = 1$:

$$\sum_{j=0}^{9} \hat{\pi}_j = 1$$

$$= \sum_{j=0}^{8} \hat{\pi}_q \cdot \frac{\sum_{i=1}^{N} \mathbb{I}(t_j^{(i)}=1)}{\sum_{i=1}^{N} \mathbb{I}(t_q^{(i)}=1)} + \hat{\pi}_q \cdot \overbrace{\frac{\sum_{i=1}^{N} \mathbb{I}(t_q^{(i)}=1)}{\sum_{i=1}^{N} \mathbb{I}(t_q^{(i)}=1)}}^{=1}$$

$$= \frac{\hat{\pi}_q}{\sum_{i=1}^{N} \mathbb{I}(t_q^{(i)}=1)} \sum_{j=0}^{9} \sum_{i=1}^{N} \mathbb{I}(t_j^{(i)}=1)$$

$$= \frac{\hat{\pi}_q \cdot N}{\sum_{i=1}^{N} \mathbb{I}(t_q^{(i)}=1)}$$

So $\hat{\pi}_q = \frac{\sum_{i=1}^{N} \mathbb{I}(t_q^{(i)}=1)}{N}$

$$\hat{\pi}_j = \frac{\sum_{i=1}^{N} \mathbb{I}(t_q^{(i)}=1)}{N} \cdot \frac{\sum_{j=1}^{N} \mathbb{I}(t_j^{(i)}=1)}{\sum_{j=1}^{N} \mathbb{I}(t_q^{(i)}=1)} = \frac{\sum_{i=1}^{N} \mathbb{I}(t_j^{(i)}=1)}{N}$$

$\forall j \in \{0\text{-}8\}$

$\forall j \in \{0\text{-}9\}$: $\hat{\pi}_j = \frac{\sum_{i=1}^{N} \mathbb{I}(t_j^{(i)}=1)}{N}$

---

② $\sum_{j=1}^{784} \log(p(x_j \mid c, \theta_{jc}))$

$$= \sum_{i=1}^{N} \sum_{j=1}^{784} \sum_{c=1} t_c^{(i)} \left[\log(\theta_{jc}^{x_j^{(i)}}) + \log((1-\theta_{jc})^{1-x_j^{(i)}}) \right]$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{784} \sum_{c=1} t_c^{(i)} \left[x_j^{(i)} \log(\theta_{jc}) + (1-x_j^{(i)}) \log(1-\theta_{jc}) \right]$$

$$\frac{\partial (2)}{\partial \theta_{jc}} = \sum_{i=1}^{N} t_c^{(i)} \left(\frac{x_j^{(i)}}{\theta_{jc}} + \frac{-(1-x_j^{(i)})}{1-\theta_{jc}} \right)$$

$$= \sum_{i=1}^{N} t_c^{(i)} x_j^{(i)} - t_c^{(i)} x_j^{(i)} \theta_{jc} - t_c^{(i)} \theta_{jc} + t_c^{(i)} x_j^{(i)} \theta_{jc}$$

$$= 0$$

$$\sum_{i=1}^{N} t_c^{(i)} x_j^{(i)} = \theta_{jc} \sum_{i=1}^{N} t_c^{(i)}$$

So $$\hat{\theta}_{jc} = \frac{\sum_{i=1}^{N} t_c^{(i)} x_j^{(i)}}{\sum_{i=1}^{N} t_c^{(i)}} = \frac{\sum_{i=1}^{N} \mathbb{I}(x_j^{(i)}=1 \wedge t_c^{(i)}=1)}{\sum_{i=1}^{N} \mathbb{I}(t_c^{(i)}=1)}$$

---

b

$$P(t|x,\theta,\pi) = \frac{P(t_c=1|\pi) P(x|t,\theta,\pi)}{\sum_{k=0}^{9} P(t_k=1) P(x|t,\theta,\pi)}$$

$$= \frac{P(t_c=\pi) \prod_{j=1}^{784} P(x_j|\theta_{jc},t_c)}{\sum_{k=0}^{9} P(t_k=1) \prod_{j=1}^{784} P(x_j|\theta_{jc}, c=k)}$$

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

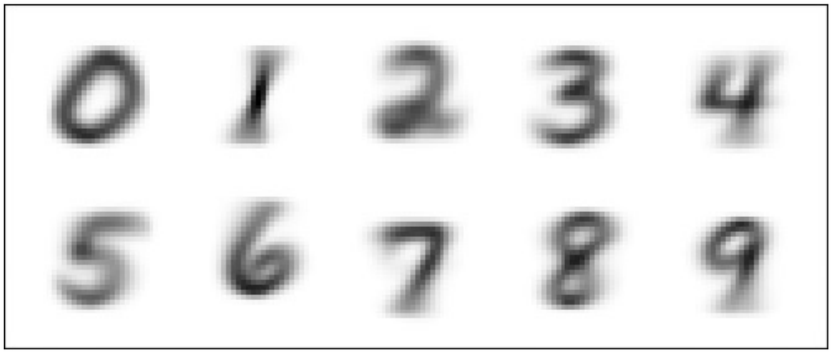
$$\log p(t|x,\theta,\pi) = \log \pi_c + \sum_{j=1}^{784} x_j \log \theta_{jc} + \sum_{j=1}^{784} (1-x_j) \log(1-\theta_{jc})$$

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

c

Undefined, since some eles in $\hat{\Theta}$ are 0 and $\log(\hat{\theta}_{jc})$ is undefined, so $\hat{\theta}_{jc} = 0$

d



e

$$P(D|\theta) = \frac{\theta^{3-1}(1-\theta)^{3-1}}{B(3,3)} = \frac{\theta^2(1-\theta)^2}{B(3,3)} \propto \theta^2(1-\theta)^2$$

by Bayes: $p(\theta|D) \propto p(\theta)p(D|\theta)$

for θ_{jc}:

$$\prod_{i=1}^{N} \mathbb{I}(t_c^{(i)}=1) \cdot \theta_{jc}^{x_j^{(i)}} (1-\theta_{jc})^{1-x_j^{(i)}} \cdot \theta_{jc}^2 (1-\theta_{jc})^2$$

$$= \theta_{jc}^{\sum_{i=N}^{N} \mathbb{I}\{c^{(i)}=c\} x_j^{(i)}} (1-\theta_{jc})^{\sum_{i=N}^{N} \mathbb{I}\{c^{(i)}=c\} - \sum_{i=N}^{N} \mathbb{I}\{c^{(i)}=c\} x_j^{(i)}} \theta_{jc}^2 \cdot (1-\theta_{jc})^2$$

$$P(\theta_{jc}|D) = \theta_{jc}^{\sum_{i=N}^{N} \mathbb{I}\{c^{(i)}=c\} x_j^{(i)} + 2} (1-\theta)^{\sum_{i=N}^{N} \mathbb{I}\{c^{(i)}=c\} - \sum_{i=N}^{N} \mathbb{I}\{c^{(i)}=c\} x_j^{(i)} + 2}$$

$$\frac{\partial \log(\theta_{jc}|D)}{\partial \log \theta_{jc}} = \left(\sum_{i=N}^{N} \mathbb{I}\{c^{(i)}=c\} x_j^{(i)} + 2\right)(1-\theta_{cj}) - \theta_{cj}\left(\sum_{i=N}^{N} \mathbb{I}\{c^{(i)}=c\} - \sum_{i=N}^{N} \mathbb{I}\{c^{(i)}=c\} x_j^{(i)} + 2\right) = 0$$

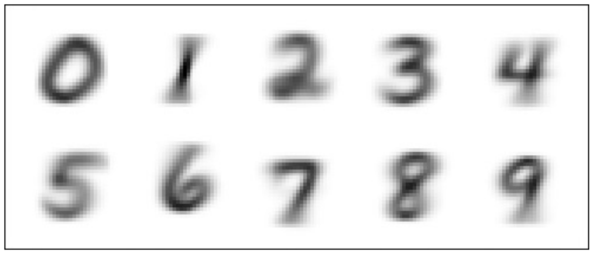
$$N+2 = \left(\sum_{i=N}^{N} \mathbb{I}\{c^{(i)}=c\} + 4\right)\theta$$

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

f

```
Average log-likelihood for MAP is  -3.3570631378602904
Training accuracy for MAP is  0.8352166666666667
Test accuracy for MAP is  0.816
```

g



3

a

By Bayes: $P(\theta|D) = \frac{P(D|\theta)\cdot P(\theta)}{P(D)} \propto P(D|\theta)\cdot P(\theta)$

Because independecy: $P(D|\theta) = \prod_{i=1}^{N} P(x^{(i)}|\theta)$

$$= \prod_{i=1}^{N}\prod_{j=1}^{k} \theta_j^{x_j^{(i)}}$$

So $P(\theta|D) \propto \prod_{i=1}^{N}\prod_{j=1}^{k} \theta_j^{x_j^{(i)}} \cdot \prod_{j=1}^{k} \theta_j^{\alpha_j - 1} = \prod_{j=1}^{k} \theta_j^{N_j + \alpha_j - 1}$

Dirichlet distribution is a conjugate prior for categorial distribution

b

$$\hat{\theta}_{map} = \underset{\theta}{\operatorname{argmax}}\, p(\theta|D)$$

$$= \underset{\theta}{\operatorname{argmax}}\, p(\theta)p(D|\theta)$$

$$= \underset{\theta}{\operatorname{argmax}}\, \log(p(\theta)\cdot p(D|\theta))$$

$$l(\theta) = \log(P(\theta)\cdot P(D|\theta)) = C + (N_1 + a_1 - 1)\cdot\log(\theta_1) + \cdots + (N_k + a_k - 1)\cdot\log(\theta_k)$$

$$\sum_{i=1}^{k}\theta_i = 1 \Rightarrow \theta_k = 1 - \sum_{i=1}^{k-1}\theta_i$$

$$l(\theta) = C + \left(\sum_{i=1}^{k-1}(N_i + a_i - 1)\log(\theta_i)\right) + (N_k + a_k - 1)\cdot\log\left(1 - \sum_{i=1}^{k-1}\theta_i\right)$$

$$\frac{\partial l(\theta)}{\partial \theta_i} = \frac{N_i + a_i - 1}{\theta_i} - \frac{N_k + a_k - 1}{1 - \sum_{i=1}^{k-1}\theta_i}$$

$$= \frac{N_i + a_i - 1}{\theta_i} - \frac{N_k + a_k - 1}{\theta_k} = 0$$

$$\frac{\theta_i}{\theta_k} = \frac{N_i + \alpha_i - 1}{N_k + \alpha_k - 1}$$

Let $\frac{\hat{\theta}_i}{\hat{\theta}_k}$ is maximized by l

since $\frac{\hat{\theta}_1}{\hat{\theta}_k} + \frac{\hat{\theta}_2}{\hat{\theta}_k} + \cdots + \frac{\hat{\theta}_k}{\hat{\theta}_k} = \frac{\sum_{i=1}^{k} \hat{\theta}_i}{\hat{\theta}_k} = \frac{1}{\hat{\theta}_k}$

$$\hat{\theta}_k = \frac{N_k + \alpha_k - 1}{\left(\sum_{i=1}^{k} N_i - \alpha_i\right) - k}$$

for $j \in \{1, 2, \cdots k\}$ $\quad \hat{\theta}_{map_j} = \frac{N_j + \alpha_j - 1}{\left(\sum_{i=1}^{k} N_i - \alpha_i\right) - k}$

C

since $P(X^{(N+1)} | D) = \int P(X^{(N+1)} | \theta) P(\theta | D) \, d\theta$

Let $X_k^{(N+1)} = 1$

$$\int P(X^{(N+1)} | \theta) \cdot P(\theta | D) \, d\theta$$

$$= \int \theta_k^{X_k^{(N+1)}} P(\theta | D) \, d\theta$$

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

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

$$= \frac{N_k + \alpha_k}{\sum_{i=1}^{k} N_i + \alpha_i}$$

4

a

```
Train average conditional log-likelihood: -0.12462443666863014.
Test average conditional log-likelihood: -0.1966732032552554.
```

b

```
Train accuracy:  0.9814285714285714.
Test accuracy: 0.97275.
```

c