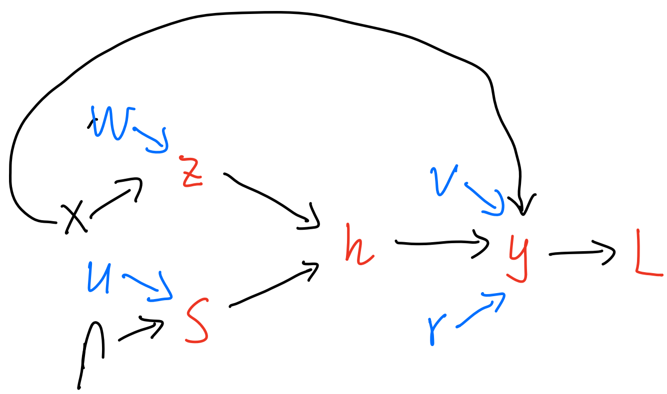
Q1:

a>



$x: N\times 1$ $z: M\times 1$

$w: M\times N$ $s: M\times 1$

$\eta: N\times 1$ $u: M\times N$

$h: M\times 1$

$v: M\times 1$ $r: N\times 1$

b> $\bar{L} = 1$

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

$\bar{v} = \bar{y}\frac{dy}{dv} = \bar{y}\cdot h$

$\bar{r} = \bar{y}\frac{dy}{dr} = \bar{y}x$

$\bar{h} = \bar{y}\frac{dy}{dh} = \bar{y}v$

$\bar{z} = \bar{h}\frac{dh}{dz} = \bar{h}\,\sigma(s)^T$

$\bar{s} = \bar{h}\frac{dh}{ds} = \bar{h}\cdot z \odot \sigma'(s)$

$\bar{u} = \bar{s}\frac{ds}{du} = \bar{s}\eta$

$\bar{\eta} = \bar{s}\frac{ds}{d\eta} = u^T\bar{s}$

$\bar{w} = \bar{z}\frac{dz}{dw} = \bar{z}x$

$\bar{x} = \bar{y}\frac{dy}{dx} + \bar{z}\frac{dz}{dx}$
$= \bar{y}r + w^T\bar{z}$

Q2 a> will derive θ_{jc};

consider the set of sample $I = \{(x^i, t^i) \mid t^i = c\}$

then by MLE, likelihood $= \theta_{jc}^{\sum_i x_{ij}} (1-\theta_{jc})^{\sum_i (1-x_{ij})}$

log likelihood is: $\sum_i x_{ij}\log(\theta_{jc}) + \sum_i (1-x_{ij})\log(1-\theta_{jc})$

where it's derivative equals to zero:

$$\frac{\sum_i x_{ij}}{\theta_{jc}} - \frac{\sum_i 1-x_{ij}}{1-\theta_{jc}} = 0 \Rightarrow \theta_{jc} = \frac{\sum_i x_{ij}}{\sum 1} = \frac{\#\text{ sample in } c \text{ with } j=1}{\#\text{ sample}}$$

i # in class C

consider π_i; rewrite $\pi_9 = 1 - \sum_{i=0}^{8} \pi_i$

thus likelihood is: $\prod_{i=0}^{8} \pi_i^{t_i} \cdot \left(1 - \sum_{i=0}^{8} \pi_i\right)^{t_9}$

log-likelihood is: $\sum_{i=0}^{8} t_i \log \pi_i + t_9 \left(1 - \sum_{i=0}^{8} \pi_i\right)$

differentiate with respect to π_i:

$$\frac{t_i}{\pi_i} - \frac{t_9}{1 - \sum_{i=0}^{8} \pi_i} = \frac{t_i}{\pi_i} - \frac{t_9}{\pi_9} = 0$$

$$\pi_i = \pi_9 \cdot \frac{t_i}{t_9}; \qquad \Rightarrow \qquad \sum_{i=0}^{9} \pi_i = \left(\frac{\sum_{i=0}^{9} t_i}{t_9}\right) \pi_9 = 1$$

then $\pi_9 = \frac{t_9}{\sum_{i=0}^{9} t_i}$ and $\pi_j = \frac{t_9}{\sum_{i=0}^{9} t_i} \cdot \frac{t_j}{t_9} = \frac{t_j}{\sum_{i=0}^{9} t_i}$

$$= \frac{\#\text{ of target} = j}{\#\text{ of total targets (samples)}}$$

b> $\log(p(t \mid x, \theta, \pi))$

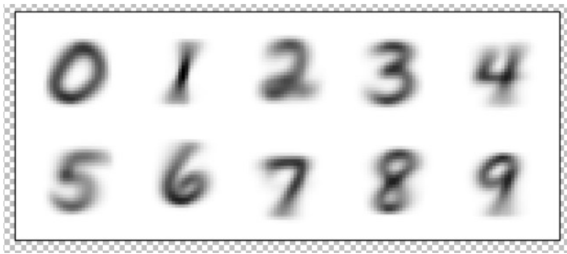
$$= \log\left(\frac{p(t \mid \pi) \prod_{j=1}^{784} \theta_{jt}^{x_j} (1 - \theta_{jt})^{(1-x_j)}}{\sum_{t'=0}^{9} p(t' \mid \pi) \prod_{j=1}^{784} \theta_{jt'}^{x_j} (1 - \theta_{jt'})^{(1-x_j)}}\right)$$

$$= \log(\pi_t) + \sum_{i=1}^{784} \left(x_j \log(\theta_{jt}) + (1 - x_j) \log(1 - \theta_{jt})\right)$$

$$- \sum_{t'=0}^{9} \left(\log(\pi_t) + \sum_{j=1}^{784} \left(x_j \log(\theta_{jt'}) + (1 - x_j) \log(1 - \theta_{jt'})\right)\right)$$

c> the avg-log-likelihood is nan, due to some probability results in zero, leading to $\log(0) = \inf$;

d>



e> MAP: let $I=\{(x^i,t^i)\mid t^i=c\}$; find θ_{jc};

log-function :

$$\log(p(\theta)) + \log(p(D|\theta)) \quad \text{(ignoring constant}$$

$$= \log\left(\theta_{jc}^2(1-\theta_{jc})^2\right) + \sum_i x_{ij}\log(\theta_{jc}) + \sum_i (1-x_{ij})\log(1-\theta_{jc})$$

$$= \left(2+\sum_i x_{ij}\right)\log(\theta_{jc}) + \left(2+\sum_i(1-x_{ij})\right)\log(1-\theta_{jc})$$

now take derivative w.r.t θ_{jc}

$$\Rightarrow \frac{2+\sum_i x_{ij}}{\theta_{jc}} - \frac{2+\sum_i(1-x_{ij})}{1-\theta_{jc}} = 0$$

$$\Rightarrow 2-2\theta_{jc}+\sum_i x_{ij} - \theta_{jc}\sum_i x_{ij} = 2\theta_{jc}+\theta_{jc}\sum_i(1-x_{ij})$$

$$-4\theta_{jc} - \sum_i(1-x_{ij}) - \sum_i x_{ij}\theta_{jc} = -2-\sum_i x_{ij}$$

then $\theta_{jc} = \dfrac{2+\sum_i x_{ij}}{4+\sum_i 1} = \dfrac{2+\#\text{ of } x^i \text{ in class } c \text{ with } j}{4+\#\text{ of class } c \text{ sample}}$

π_j: prior: $p(\pi_j)=\frac{1}{10}$

$$MAP(\pi_j) = \log(p(\pi)) + \log(p(t|\pi))$$

$$= \log\left(\frac{1}{10}\right) + \log\left(\prod_{i=0}^{8}\pi_i^{t_i}\cdot\left(1-\sum_{i=0}^{8}\pi_i\right)^{t_9}\right)$$

$$= \log\left(\frac{1}{10}\right) + \sum_{i=0}^{8} t_i\log(\pi_i) + t_9\log\left(1-\sum_{i=0}^{8}\pi_i\right)$$

take derivative w.r.t π_j :

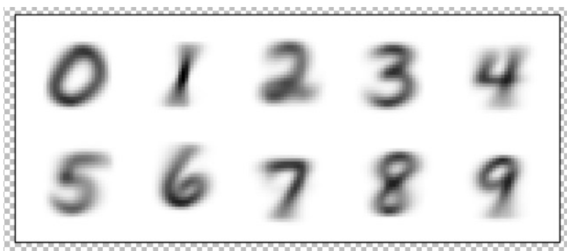
$$\frac{d}{d\pi_j} = 0 + \frac{t_j}{\pi_j} - \frac{t_9}{1-\sum_{i=0}^{8}\pi_i} = \frac{t_j}{\pi_j} - \frac{t_9}{1-\sum_{i=0}^{8}\pi_i} = 0$$

thus by part (a), $\pi_j = \dfrac{\#\text{ of target} = j}{\#\text{ of total targets (samples)}}$

f> the avg log-likelihood is -1.0578

the accuracy on training set is: 0.83521666...

on test set is: 0.816

g> 

Q3:

a> use Bayes rule:

$$p(\theta|D) \propto p(\theta)\cdot p(D|\theta) = \theta_1^{a_1-1}\cdots\theta_k^{a_k-1}\cdot\prod_{i=1}^{N}\left(\prod_{k=1}^{k}\theta_k^{x_{ik}}\right)$$

$$= \prod_{k=1}^{k}\left(\theta_k^{a_k-1}\cdot\prod_{i=1}^{N}\theta_k^{x_{ik}}\right) = \prod_{k=1}^{k}\theta_k^{\sum_i(x_{ik})+a_k-1}$$

since the posterior distribution has the same form as prior, this distribution is conjugate.

b> $\log(P(\theta)) + \log(P(D|\theta))$ let $\theta_k = 1 - \sum_{j=1}^{k-1} \theta_j$

$$= \log\left(\prod_{j=1}^{k-1} \theta_j^{a_j - 1} \cdot \left(1 - \sum_{j=1}^{k-1} \theta_j\right)^{a_k - 1}\right) +$$

$$\log\left(\prod_{j=1}^{k-1} \theta_j^{\sum_i x_{ij}} \cdot \left(1 - \sum_{j=1}^{k-1} \theta_j\right)^{\sum_i x_{ik}}\right)$$

$$= \sum_{j=1}^{k-1} (a_j - 1) \log(\theta_j) + (a_k - 1) \log\left(1 - \sum_{j=1}^{k-1} \theta_j\right)$$

$$+ \sum_{j=1}^{k-1} \left(\sum_i x_{ij}\right) \log \theta_j + \left(\sum_i x_{ik}\right) \log\left(1 - \sum_{j=1}^{k-1} \theta_j\right)$$

differentiate w.r.t. θ_j

$$\Rightarrow \frac{a_j - 1}{\theta_j} + \frac{a_k - 1}{\theta_k}(-1) + \frac{\sum_i x_{ij}}{\theta_j} - \frac{\sum_i x_{ik}}{\theta_k} = 0$$

for convenience, let $\sum_i x_{ij} = N_j$ $\sum_i x_{ik} = N_k$

then $\theta_j = \dfrac{\theta_k (N_j + a_j - 1)}{N_k + a_k - 1}$

realizing $\sum_{i=1}^{k} \theta_i = 1 \Rightarrow \theta_k \left(\dfrac{\sum_{j=1}^{k} (N_j + a_j - 1)}{N_k + a_k - 1}\right) = 1$

then $\theta_k = \dfrac{N_k + a_k - 1}{N_{total} - k + \sum_{j=1}^{k} a_j}$,

$$\theta_i = \frac{\theta_k (N_j + a_j - 1)}{N_k + a_k - 1} = \frac{N_k + a_k - 1}{N_{total} - k + \sum_{j=1}^{k} a_j} \times \frac{(N_j + a_j - 1)}{N_k + a_k - 1} = \frac{N_j + a_j - 1}{N_{total} - k + \sum_{j=1}^{k} a_j}$$

c> ?

Q4

a> train log-likelihood : -0.12462

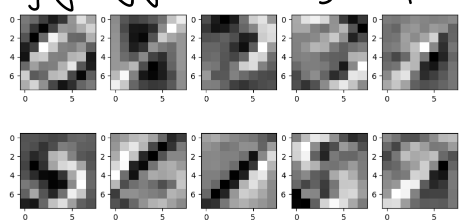
test log-likelihood : -0.19667

b> accuracy on train: 0.98142857

accuracy on test: 0.97275

c>

fig0 fig1 2 3 4



5 6 7 8 9