Q1.

(a)

$$err'_t = \frac{\sum_{i=1}^N w_i' \mathbb{I}\{h_t(x^{(i)}) \neq t^{(i)}\}}{\sum_{i=1}^N w_i'}$$

$$= \frac{\sum_{i\in E} w_i' \cdot 1 + \sum_{i\in E^c} w_i' \cdot 0}{\sum_{i=1}^N w_i'} = \frac{\sum_{i\in E} w_i'}{\sum_{i=1}^N w_i'}$$

$$= \frac{\sum_{i\in E} w_i \exp(-\alpha_t \cdot t^{(i)} \cdot h_t(x^{(i)}))}{\sum_{i=1}^N w_i \exp(-\alpha_t \cdot t^{(i)} \cdot h_t(x^{(i)}))}$$

$$= \frac{\sum_{i\in E} w_i \exp(\alpha_t)}{\sum_{i\in E} w_i \exp(\alpha_t) + \sum_{i\in E^c} w_i \exp(-\alpha_t)}$$, plug in $\alpha_t = \frac{1}{2}\log\frac{1-err_t}{err_t}$

$$= \frac{\sum_{i\in E} w_i \left(\frac{1-err_t}{err_t}\right)^{\frac{1}{2}}}{\sum_{i\in E} w_i \left(\frac{1-err_t}{err_t}\right)^{\frac{1}{2}} + \sum_{i\in E^c} w_i \left(\frac{1-err_t}{err_t}\right)^{-\frac{1}{2}}}$$

$$= \frac{\sum_{i\in E} w_i \cdot \frac{1-err_t}{err_t}}{\sum_{i\in E} w_i \frac{1-err_t}{err_t} + \sum_{i\in E^c} w_i}$$

, since $\frac{1-err_t}{err_t} = \frac{\sum_{i\in E^c} w_i / \sum^N w_i}{\sum_{i\in E} w_i / \sum^N w_i} = \frac{\sum_{i\in E^c} w_i}{\sum_{i\in E} w_i}$

we know $\sum_{i\in E^c} w_i = \frac{1-err_t}{err_t} \cdot \sum_{i\in E} w_i$

so

$$err'_t = \frac{\sum_{i\in E} w_i \cdot \frac{1-err_t}{err_t}}{\sum_{i\in E} w_i \cdot \frac{1-err_t}{err_t} + \sum_{i\in E} w_i \frac{1-err_t}{err_t}} = \frac{1}{2}.$$

Interpretation: In the new iteration, the algorithm always adjust the weight s.t. classification error becomes 1/2. That is, we increase weight on the datapoints we got wrong before. Also since $err' \le \frac{1}{2}$, $\alpha \ge 0$. α decreases as err increases, so the classifiers w/ smaller error does more contribution.

(b) $$\frac{w_i \cdot \exp(2\alpha_t \mathbb{I}\{h_t(x^{(i)}) \neq t^{(i)}\})}{w_i \cdot \exp(-\alpha_t \cdot t^{(i)} \cdot h_t(x^{(i)}))}$$

$$= \exp\left[\alpha_t \cdot (1 - h_t(x^{(i)}) \cdot t^{(i)}) - (-\alpha_t \cdot t^{(i)} \cdot h_t(x^{(i)}))\right]$$

$$= \exp\left[\alpha_t - \alpha_t \cdot h_t(x^{(i)}) \cdot t^{(i)} + \alpha_t \cdot t^{(i)} \cdot h_t(x^{(i)})\right]$$

$$= \exp(\alpha_t)$$

so the constant factor is e^{α_t}, i.e. $\sqrt{\frac{1-err_t}{err_t}}$

Q2

(a)

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

$$= \sum_{i=1}^{N} \log(p(x^{(i)} \mid t^{(i)}, \theta, \pi)\, p(t^{(i)} \mid \theta, \pi))$$

$$= \sum_{i=1}^{N} \log(p(t^{(i)} \mid \pi) \prod_{j=1}^{784} p(x_j^{(i)} \mid t^{(i)}, \theta_{jc^{(i)}}))$$

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

$$= \sum_{i=1}^{N} \sum_{j=0}^{9} \log \pi_j^{t_j^{(i)}} + \sum_{i=1}^{N} \sum_{j=1}^{784} \log \left[\theta_{jc^{(i)}}^{x_j^{(i)}} \left(1-\theta_{jc^{(i)}}\right)^{1-x_j^{(i)}} \right]$$

$$= \sum_{i=1}^{N} \left[\sum_{j=0}^{8} t_j^{(i)} \log \pi_j + t_q^{(i)} \log\left(1-\sum_{j=0}^{8} \log \pi_j\right) \right] +$$

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

For $j = 0, 1, \cdots, 8$:

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

$$= \sum_{i=1}^{N} t_j^{(i)} \cdot \frac{1}{\pi_j} - t_q^{(i)} \frac{1}{\pi_q}$$

set it to 0:

$$\sum_{i=1}^{N} t_j^{(i)} \frac{1}{\pi_j} = \sum_{i=1}^{N} t_q^{(i)} \frac{1}{\pi_q}$$

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

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

$$\left[\sum_{j=0}^{8} \frac{\sum_{i=1}^{N} t_j^{(i)}}{\sum_{i=1}^{N} t_q^{(i)}} \hat{\pi}_q \right] + \hat{\pi}_q = 1$$

$$\left(\frac{\sum_{i=1}^{N} \sum_{j=0}^{9} t_j^{(i)}}{\sum_{i=1}^{N} t_q^{(i)}} \right) \cdot \hat{\pi}_q = 1 \quad , \quad \text{by} \sum_{j=0}^{9} t_j^{(i)} = 1,$$

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

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

$$\text{so } \hat{\pi}_j = \frac{\sum_{i=1}^{N} t_j}{N} \quad , \quad j = 0, \cdots, 9.$$

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

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

$$\text{set } \frac{\partial \ell}{\partial \theta_{jc}} = 0,$$

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

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

$$= \log \frac{p(c \mid \pi) p(x \mid c, \theta)}{\sum_{c'} p(c' \mid \pi) p(x \mid c', \theta_{jc'})}$$

$$= \log \frac{\pi_c \cdot \prod_{j=1}^{784} p(x_j \mid c, \theta_{jc})}{\sum_{c'} \pi_{c'} \cdot \prod_{j=1}^{784} p(x_j \mid c', \theta_{jc'})}$$

$$= \log \frac{\pi_c \cdot \prod_{j=1}^{784} \theta_{jc}^{x_j} (1-\theta_{jc})^{(1-x_j)}}{\sum_{c'} \pi_{c'} \cdot \prod_{j=1}^{784} \theta_{jc'}^{x_j} (1-\theta_{jc'})^{(1-x_j)}}$$

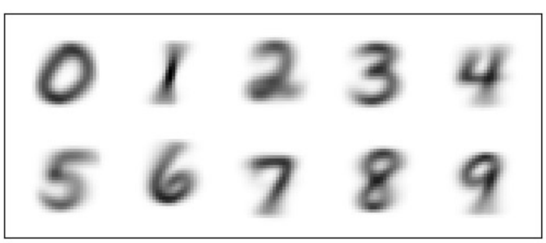
$$= \log \pi_c + \sum_{j=1}^{784} \left[x_j \log \theta_{jc} + (1-x_j) \log (1-\theta_{jc}) \right]$$
$$- \log \left[\sum_{c'=0}^{9} \pi_{c'} \cdot \prod_{j=1}^{784} \theta_{jc'}^{x_j} (1-\theta_{jc'})^{(1-x_j)} \right]$$

(c)

```
/Users/yukino/Documents/UToronto/CSC311/hw3/naive_bayes.py:126: RuntimeWarning: divide by zero encountered in log
  temp = np.log(pi) + images @ np.log(theta) + (ones_mat1 - images) @ np.log(ones_mat2 - theta)
/Users/yukino/Documents/UToronto/CSC311/hw3/naive_bayes.py:126: RuntimeWarning: invalid value encountered in matmul
  temp = np.log(pi) + images @ np.log(theta) + (ones_mat1 - images) @ np.log(ones_mat2 - theta)
Average log-likelihood for MLE is  nan
```

When the count of datapoints for some class equals to 0,
we may encounter log 0 and divide by 0 in our computation.

(d)



(e) $\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ p(\theta \mid x)$

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

$= \underset{\theta}{\operatorname{argmax}}\ p(x \mid \theta)\, p(\theta)$

$= \underset{\theta}{\operatorname{argmax}}\ \log p(x \mid \theta) + \log p(\theta)$

$\frac{\partial}{\partial \theta_{jc}} \left[\log p(x \mid \theta) + \log p(\theta) \right]$, by $\theta \sim \text{Beta}(3,3)$

$= \frac{\partial}{\partial \theta_{jc}} \left[\log \prod_{i=1}^{N} p(c^{(i)} \mid \pi) \prod_{j=1}^{784} \theta_{jc}^{x_j} (1-\theta_{jc})^{(1-x_j)} + \log \frac{\Gamma(6)}{\Gamma(3)\Gamma(3)} \theta^2 (1-\theta)^2 \right.$
$\left. + \sum_{i=1}^{N} \sum_{k=0}^{9} t_k^{(i)} \log \pi_k \right]$

$= \sum_{i=1}^{N} \mathbb{I}\{c^{(i)} = c,\ x_j^{(i)} = 1\} \cdot \frac{1}{\theta_{jc}} - \sum_{i=1}^{N} \mathbb{I}\{c^{(i)} = c,\ x_j^{(i)} = 0\} \frac{1}{1-\theta_{jc}}$
$+ \frac{2}{\theta_{jc}} - \frac{2}{1-\theta_{jc}}$ (ignoring $\frac{\Gamma(6)}{\Gamma(3)\Gamma(3)}$)

$= \frac{1}{\theta_{jc}} \left[\sum_{i=1}^{N} \mathbb{I}\{c^{(i)} = c,\ x_j^{(i)} = 1\} + 2 \right] - \frac{1}{1-\theta_{jc}} \left[\sum_{i=1}^{N} \mathbb{I}\{c^{(i)} = c,\ x_j^{(i)} = 0\} + 2 \right]$

set it to 0:

$(1-\hat{\theta}_{jc}) \left[\sum_{i=1}^{N} \mathbb{I}\{c^{(i)} = c,\ x_j^{(i)} = 1\} + 2 \right] = \hat{\theta}_{jc} \left[\sum_{i=1}^{N} \mathbb{I}\{c^{(i)} = c,\ x_j^{(i)} = 0\} + 2 \right]$

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

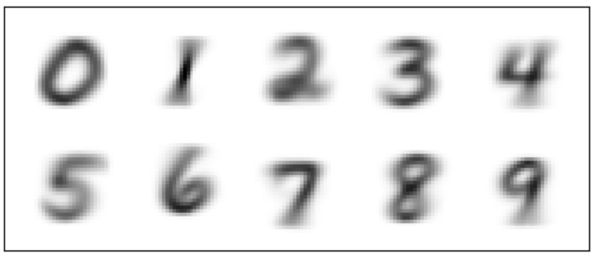
The partial derivative w.r.t π_j is same as (a), so

$$\hat{\pi}_j = \frac{\sum_{n=1}^{N} t_j}{N}$$

(f)

```
Average log-likelihood for MAP is  -3.3570631378602847
Training accuracy for MAP is  0.8352166666666667
Test accuracy for MAP is  0.816
```

(g)



3.

(a) True

(b) False

(c)