csc412 assignment 2

1. (a)

$$P(\theta|\boldsymbol{x}, c) = P(\boldsymbol{x}|c, \theta) \cdot P(\theta)$$

$$logP(\theta|\boldsymbol{x}, c) = log(P(\boldsymbol{x}|c, \theta)) + log(P(\theta))$$

$$= log(\prod_{d=1}^{784} p(x_d|c, \theta_{cd})) + log(P(\theta))$$

$$= \sum_{d=1}^{784} log(p(x_d|c, \theta_{cd})) + log(P(\theta))$$

$$= \sum_d log(\theta_{cd}^{x_d}(1-\theta_{cd})^{(1-x_d)}) + log(\theta_{cd}(1-\theta_{cd}))$$

$$= \sum_d [x_d log(\theta_{cd}) + (1-x_d)log(1-\theta_{cd}))] + log(\theta_{cd}) + log(1-\theta_{cd})$$

$$\frac{\partial}{\partial \theta_{cd}} logP(\theta|\boldsymbol{x}, c) = \sum_d [\frac{x_d}{\theta_{cd}} + \frac{1-x_d}{\theta_{cd}-1}] + \frac{1}{\theta_{cd}} + \frac{1}{\theta_{cd}-1} = 0$$

$$= \sum \frac{x_d}{\theta_{cd}} + \sum \frac{1-x_d}{\theta_{cd}-1} + \frac{1}{\theta_{cd}} + \frac{1}{\theta_{cd}-1}$$

$$\frac{1}{\theta_{cd}} \sum x_d + \frac{1}{\theta_{cd}-1} \sum (1-x_d) + \frac{1}{\theta_{cd}} + \frac{1}{\theta_{cd}-1}$$

$$= \frac{(\theta_{cd}-1)\sum x_d + \theta_{cd}(N - \sum x_d) + \theta_{cd} - 1 + \theta_{cd}}{\theta_{cd}(\theta_{cd}-1)} \text{where N is number of image}$$

$$(\theta_{cd}-1)\sum x_d + \theta_{cd}(N - \sum x_d) + \theta_{cd} - 1 + \theta_{cd} = 0$$

$$\theta_{cd}\sum x_d - \sum x_d + \theta_{cd}(N - \sum x_d) + \theta_{cd} - 1 + \theta_{cd} = 0$$

$$\theta_{cd}(\sum x + d + N - \sum x_d + 1 + 1) - \sum x_d - 1 = 0$$

$$\theta_{cd} = \frac{\sum x_d + 1}{N + 2}$$

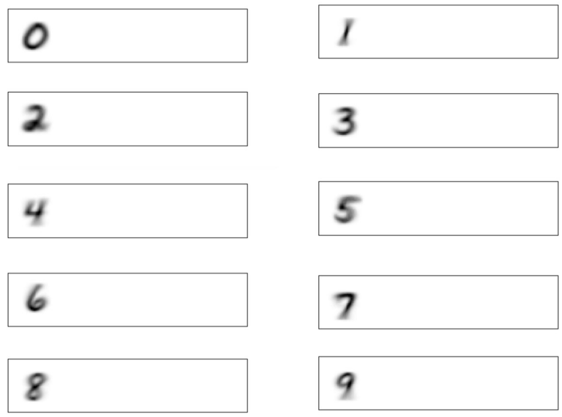
(b) The $\boldsymbol{\theta}$ are as follow:(Please refer to figure 1)

(c)

$$logp(c|x, \theta, \pi) = log\big(\frac{P(c, x)|\pi, \theta}{\sum_k P(c = k, x|\pi, \theta)}\big)$$

$$= logP(c, x|\pi, \theta) - log\sum_k P(c = k, x|\pi, \theta)$$

Figure 1: θ for 1b



$$= log\big[p(c|\pi)\prod_{d=1}^{784} p(x_d|c,\theta_{cd})\big] - log\sum_k P(c=k,x|\pi,\theta)$$

$$= logp(c|\pi) + \sum_d logP(x_d|c,\theta_{cd}) - log\sum_k P(c=k,x|\pi,\theta)$$

$$= log\pi_c + \sum_d log[\theta_{cd}^{x_d}(1-\theta_{cd})^{(1-x_d)}] - log\sum_k P(c=k,x|\pi,\theta)$$

$$= log\pi_c + \sum_d log\theta_{cd}^{x_d} + \sum_d log(1-\theta_{cd}^{(1-x_d)}) - log\sum_k P(c=k,x|\pi,\theta)$$

$$= log\pi_c + \sum_d x_d log\theta_{cd} + \sum_d (1-x_d)log(1-\theta_{cd}) - log\sum_k \pi_k \prod_d \theta_{cd}^{x_d}(1-\theta_{cd})^{(1-x_d)}$$

(d) Log likelihood for training set : -171.348616694
Log likelihood for test set :-171.743013392
Accuracy for training set: 0.8404
Accuracy for test set: 0.8414

2. (a) True, this is the assumption of naive bayes.

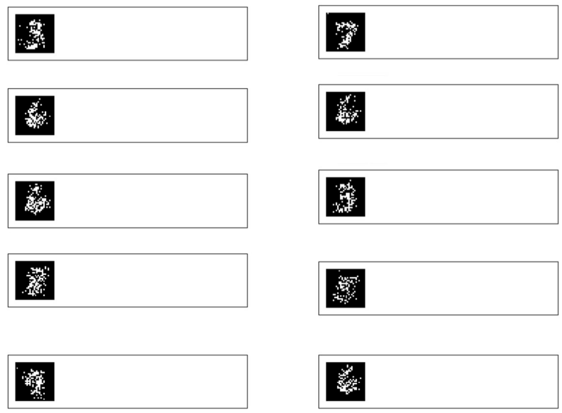
(b) False, no assumption regarding marginal independence. Conditionally independent does not imply marginal independent.

(c) The random image samples are as follow: (Please refer to figure 2)

(d)

$$P(x_b|x_t,\theta,\pi) = \sum_k P(x_b, c=k|x_t,\theta,\pi)$$

Figure 2: Samples for 2c



$$\sum_k P(x_b|c=k, x_t, \theta, \pi) \cdot P(c=k|x_t, \theta, \pi)$$

$$\sum_k P(x_b|c=k, x_t, \theta, \pi) \cdot \frac{p(c=k, x_t|\theta, \pi)}{p(x_t|\theta, \pi)}$$

$$\sum_k P(x_b|c=k, x_t, \theta, \pi) \cdot \frac{p(c=k, x_t|\theta, \pi) \cdot p(c=k|\pi, \theta)}{\sum_{k'} p(x_t, c=k'|\theta, \pi)}$$

$$\frac{1}{10} \sum_k P(x_b|c=k, x_t, \theta, \pi) \cdot \frac{p(c=k, x_t|\theta, \pi)}{\sum_{k'} p(k'|\pi) \prod_{d=1}^{392} p(x_d|k', \theta_{k'd})}$$

$$= \sum_k P(x_b|c=k, x_t, \theta, \pi) \cdot \frac{p(c=k, x_t|\theta, \pi)}{\sum_{k'} \prod_{d=1}^{392} p(x_d|k', \theta_{k'd})}$$

$$= \sum_k P(x_b|c=k, x_t, \theta, \pi) \cdot \frac{\prod_{d=1}^{392} p(x_d|k, \theta)}{\sum_{k'} \prod_{d=1}^{392} p(x_d|k', \theta_{k'd})}$$

where $p(x_d|k, \theta_{cd}) = Ber(x_d|\theta_{cd})$

(e)

$$p(x_{i \in bottom}|x_t, \theta, \pi) \sum_k P(x_i, c=k|x_t, \theta, \pi)$$

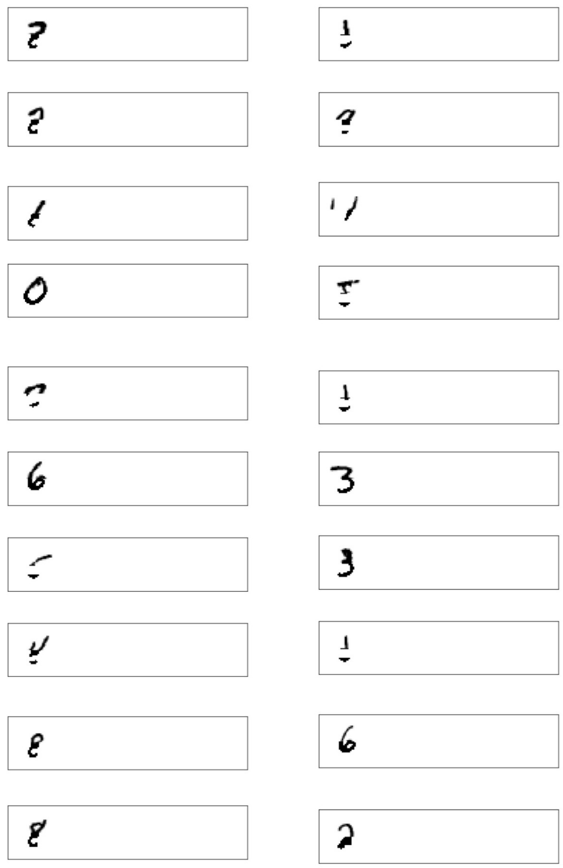
$$\sum_k P(x_i|c=k, x_t, \theta, \pi) \cdot P(c=k|x_t, \theta, \pi)$$

$$\sum_k P(x_i|c=k, x_t, \theta, \pi) \cdot \frac{p(c=k, x_t|\theta, \pi)}{p(x_t|\theta, \pi)}$$

$$\sum_k P(x_i|c=k, x_t, \theta, \pi) \cdot \frac{\prod_{d=1}^{392} p(x_d|k,\theta)}{\sum_{k'} \prod_{d=1}^{392} p(x_d|k', \theta_{k'd})}$$

(f) The augmented images are as follow:(Please refer to figure 3)

Figure 3: Images for 2f



3. (a) 784 weights for each class, 10 classes in total hence 784 x 10 = 7840

(b) 784 x 9 = 7056. If we know weights for 9 of the classes, then we know the probability of the 9 classes. Hence we could calculate the probability for the 10th class since all probability should add up to 1.

(c) Calculate $\frac{\partial logp(c|x,w)}{\partial w_i}$.

$$\frac{\partial logp(c|x,w)}{\partial w_i} = \frac{\partial}{\partial w_i}\left[log(exp(w_c^T x)) - log(\sum_{c'=0}^{9} exp(w_{c'}^T x)\right]$$

If $i \neq c$:

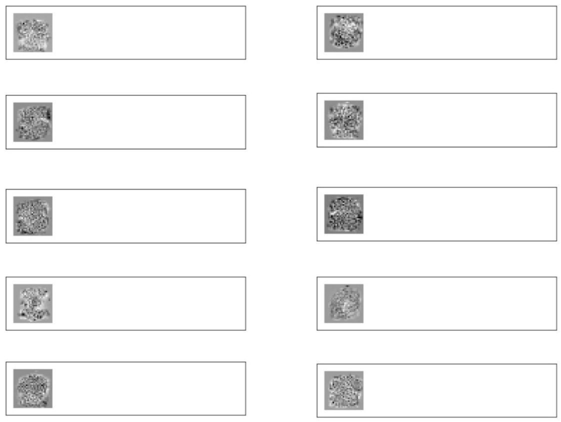
$$\frac{\partial logp(c|x,w)}{\partial w_i} = -\frac{1}{\sum_{c'} exp(w_{c'}^T x)} \cdot exp(w_i^T x) \cdot x = -xp(c|x,w)$$

If $i = c$:

$$\frac{\partial logp(c|x,w)}{\partial w_i} = \frac{1}{exp(w_c^T x)} \cdot exp(w_c^T x)x - \frac{1}{\sum_{c'} exp(w_{c'}^T x)} \cdot exp(w_i^T x) \cdot x$$
$$= x - \frac{1}{\sum_{c'} exp(w_{c'}^T x)} \cdot exp(w_i^T x) \cdot x$$
$$= x - p(c|x,w) \cdot x$$

(d) The weights are as follow:(Please refer to figure 4)

Figure 4: Weights for 3d



(e) Train set accuracy: 1
Test set accuracy: 0.8755
Train set log likelihood: -138620.483702
Test set log likelihood: -5.21722325971e-08

4. (a) 784 weights for each class, 10 classes in total. 1 prior for each class. Hence 784K + 1K = 785K

(b) K!. Since switching any two classes does not affect anything. However, switching the order of weights inside weight matrix of one of the classes would change the prediction.

(c)

$$p(\bar{x}|\theta,\pi) = \sum_{c=1}^{K} \pi_c \prod_{d=1}^{784} \theta_{cd}^{x_d}(1-\theta_{cd})^{(1-x_d)}$$

$$logp(\bar{x}|\theta,\pi) = log[\sum_{c=1}^{K} \pi_c \prod_{d=1}^{784} \theta_{cd}^{x_d}(1-\theta_{cd})^{(1-x_d)}]$$

$$\frac{\partial logp(\bar{x}|\theta,\pi}{\partial\theta_{cd}} =$$

$$\frac{1}{\sum_{c=1}^{K}\pi_c\prod_{d=1}^{784}\theta_{cd}^{x_d}(1-\theta_{cd})^{(1-x_d)}}\cdot\pi_c\frac{d}{d\theta_{cd}}[\theta_{cd}^{x_d}(1-\theta_{cd})^{(1-x_d)}]\cdot\prod_{d'=1,d'\neq d}^{784}\theta_{cd'}^{x_{d'}}(1-\theta_{cd'})^{(1-x_{d'})}$$

$$=\frac{\pi_c[\theta_{cd}^{x_d}(1-x_d)(1-\theta_{cd})^{(-x_d)}(-1)+(1-\theta_{cd})^{(1-x_d)}(x_d)\theta_{cd}^{(x_d-1)}]\cdot\prod_{d'=1,d'\neq d}^{784}\theta_{cd'}^{x_{d'}}(1-\theta_{cd'})^{(1-x_{d'})}}{P(\bar{x}|\theta,\pi)}$$

(d) The $\boldsymbol{\theta}$ are as follow: The clusters are not as blurry as other supervised model.Hence higher level of confidence. Also categories with more variance has more class in this model.

(e) The augmented images are as follow: The result of this model look more completed than naive bayes model. In the result from naive bayes, strokes on some images are not continuous. That is, the bottom half failed to connect with the stroke on the top half. Where as using EM algorithm, the bottom half successfully connect with the top half.

Figure 5: θ for 4d

Figure 6: Images for 4e