Q1

1):

$p(y=k \mid x, \mu, \sigma)$

$$= p(x \mid y=k, \mu, \sigma) \cdot p(y) \,/\, p(x \mid \mu, \sigma)$$

$$= \frac{\left[\left(\prod_{i=1}^{D} 2\pi\sigma_i^2\right)^{-\frac{1}{2}} \cdot \exp\left(-\sum_{i=1}^{D} (x_i-\mu_{ki})^2/2\sigma_i^2\right)\right]\alpha_k}{\sum_{k=1}^{K}\left\{\left[\left(\prod_{i=1}^{D} 2\pi\sigma_i^2\right)^{-\frac{1}{2}} \cdot \exp\left(-\sum_{i=1}^{D} (x_i-\mu_{ki})^2/2\sigma_i^2\right)\right]\alpha_k\right\}}$$

2):

$L(\theta, D)$

$$= -\log\left(P(x, y \mid \theta)\right)$$

$$= -\log \prod_{n=1}^{N} P(x^{(n)}, y^{(n)} \mid \theta)$$

$$= -\sum_{n=1}^{N}\sum_{k=1}^{K} \mathbb{1}(y^{(n)}=k)\, \log\left(P(x^{(n)} \mid y^{(n)}=k, \theta)\, P(y^{(n)}=k)\right)$$

$$= -\sum_{n=1}^{N}\sum_{k=1}^{K} \mathbb{1}(y^{(n)}=k)\, \log\left[\left(\prod_{i=1}^{D} 2\pi\sigma_i^2\right)^{-\frac{1}{2}} \exp\left(-\sum_{i=1}^{D} (x_i^{(n)}-\mu_{ki})^2/2\sigma_i^2\right)\cdot \alpha_k\right]$$

$$= -\sum_{n=1}^{N}\sum_{k=1}^{K} \mathbb{1}(y^{(n)}=k)\left[\left(\frac{-1}{2}\sum_{i=1}^{D}\log(2\pi)\right)-\left(\frac{1}{2}\sum_{i=1}^{D}\log(\sigma_i^2)\right)-\left(\frac{1}{2}\sum_{i=1}^{D}\frac{(x_i^{(n)}-\mu_{ki})^2}{\sigma_i^2}\right)+\log\alpha_k\right]$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} \mathbb{1}(y^{(n)}=k)\left[\left(\frac{D}{2}\log(2\pi)\right)+\left(\frac{1}{2}\sum_{i=1}^{D}\log(\sigma_i^2)\right)+\left(\frac{1}{2}\sum_{i=1}^{D}\frac{(x_i^{(n)}-\mu_{ki})^2}{\sigma_i^2}\right)-\log\alpha_k\right]$$

3):

$\partial L(\theta, D) / \partial \mu_{ki}$

$$= \partial \sum_{n=1}^{N} \mathbb{1}(y^{(n)}=k)\, \frac{(x_i^{(n)}-\mu_{ki})^2}{2\sigma_i^2} \,/\, \partial \mu_{ki}$$

$$= \sum_{n=1}^{N} \mathbb{1}(y^{(n)}=k)\left(\frac{\mu_{ki}-x_i^{(n)}}{\sigma_i^2}\right)$$

$\partial L(\theta, D) / \partial \sigma_i^2$

$$= \partial \sum_{n=1}^{N}\sum_{k=1}^{K} \mathbb{1}(y^{(n)}=k)\left[\frac{1}{2}\log\sigma_i^2 + \frac{1}{2}\frac{(x_i^{(n)}-\mu_{ki})^2}{\sigma_i^2}\right] / \partial \sigma_i^2$$

$$= \partial \sum_{n=1}^{N}\left[\frac{1}{2}\log\sigma_i^2 + \frac{1}{2}\frac{(x_i^{(n)}-\mu_{y^{(n)}i})^2}{\sigma_i^2}\right] / \partial \sigma_i^2$$ # change from μ_{ki} to $\mu_{y^{(n)}i}$

$$= \frac{1}{2}\sum_{n=1}^{N}\left[\frac{1}{\sigma_i^2} - \frac{(x_i^{(n)}-\mu_{y^{(n)}i})^2}{(\sigma_i^2)^2}\right]$$

4):

$$\partial L(\theta, D) / \partial \mu_{ki} = 0$$

$$\Rightarrow \sum_{n=1}^{N} \mathbb{1}(y^{(n)} = k) \left(\frac{\mu_{ki} - x_i^{(n)}}{\sigma_i^2} \right) = 0$$

$$\Rightarrow \sum_{n=1}^{N} \mathbb{1}(y^{(n)} = k) \mu_{ki} = \sum_{n=1}^{N} \mathbb{1}(y^{(n)} = k) x_i^{(n)}$$

$$\Rightarrow \mu_{ki} = \sum_{n=1}^{N} \mathbb{1}(y^{(n)} = k) x_i^{(n)} \Big/ \sum_{n=1}^{N} \mathbb{1}(y^{(n)} = k)$$

$$\partial L(\theta, D) / \partial \sigma_i^2 = 0$$

$$\Rightarrow \frac{1}{2} \sum_{n=1}^{N} \left[\frac{1}{\sigma_i^2} - \frac{(x_i^{(n)} - \mu_{y^{(n)}i})^2}{(\sigma_i^2)^2} \right] = 0$$

$$\Rightarrow \sum_{n=1}^{N} \frac{1}{\sigma_i^2} = \sum_{n=1}^{N} \frac{(x_i^{(n)} - \mu_{y^{(n)}i})^2}{(\sigma_i^2)^2}$$

$$\Rightarrow N \frac{1}{\sigma_i^2} = \frac{1}{\sigma_i^4} \sum_{n=1}^{N} (x_i^{(n)} - \mu_{y^{(n)}i})^2$$

$$\Rightarrow N \sigma_i^2 = \sum_{n=1}^{N} (x_i^{(n)} - \mu_{y^{(n)}i})^2$$

$$\Rightarrow \sigma_i^2 = \frac{1}{N} \sum_{n=1}^{N} (x_i^{(n)} - \mu_{y^{(n)}i})^2$$

Q2.0 plot mean of each digit
plot

Q2.1

1): For k=1, report train, test classification accuracy

For k=15, report train, test classification accuracy

2): when encounter ties
decrement k and compute again until ties breaks.

Since we are still using K-NN with a different k, it is reasonably accurate.
Also, when k reduces to 1, there will be no tie which ensures breaking tie

3): Use 10-fold cross validation:
report k, train_accuracy, average_accuracy, test_accuracy.

Q2.2

1):

plot log of diagonals of covariances

plot

2): report average_cond_log_likelihood

$\frac{1}{N}\sum_{i=1}^{N} \log(p(y^{(i)} \mid x^{(i)}, \theta))$ for both test, train

3): classify and report accuracy on train, test

Q2.3

1). 2). No report needed.

3): plot eta of each class
plot

4): sample data. plot

average sampled data. plot