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):

$L(\theta, D)$

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

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

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

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

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

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

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):

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

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

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

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

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

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

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

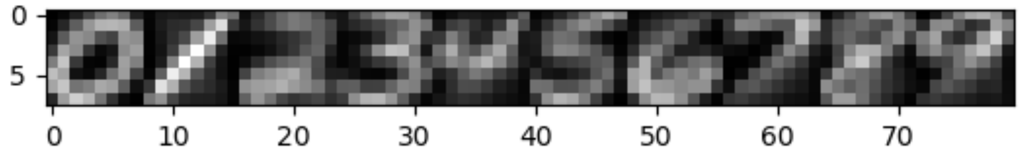
$$\Rightarrow N \frac{1}{\sigma_i^2} = \frac{1}{\sigma_i^4} \sum_{n=1}^{N} (x_i^{(n)} - \mu_{y^{(n)}i})^2$$

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

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

Q2.0

Plot of mean of each class/digit



Q2.1

1):

k= 1 | train accuracy: 1.00000 | test accuracy: 0.96875
k=15 | train accuracy: 0.96100 | test accuracy: 0.95925

2):

When encounter ties, I decrement k and compute again until the tie break.
Since I am still using K-NN with a different k. Thus ensures accuracy.
Since when k reduces to 1, there can be no tie. Thus ensures breaking of ties.

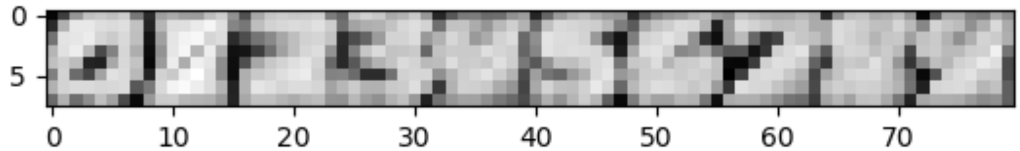
3):

optimal_k: 3
classification accuracy for train set: 0.981429
average accuracy across folds: 0.967000
classification accuracy for test set: 0.972750

Q2.2

1):

Plot of log of diagonals of covariances



2):

average conditional log likelihood for train set: -0.124624
average conditional log likelihood for test set: -0.196673

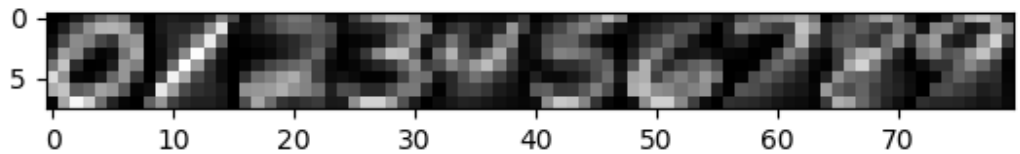
3):

classification accuracy for train set: 0.981429
classification accuracy for test set: 0.972750

Q2.3

3):

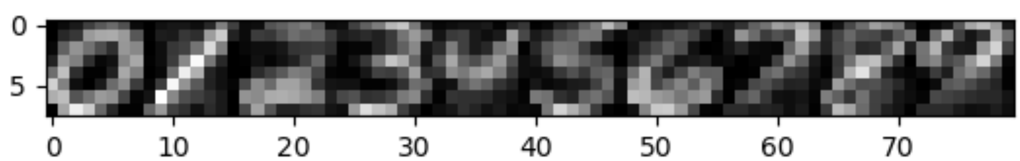
Plot of eta of each class/digit



4):

Plot of a sampled data

Plot of average of 100 sampled data



5):

average conditional log likelihood for train set: -0.943754
average conditional log likelihood for test set: -0.987270

6):

classification accuracy for train set: 0.774143
classification accuracy for test set: 0.764250

Q2.4

K-NN

classification accuracy for train set: 0.986571
classification accuracy for test set: 0.969750

Gaussian

classification accuracy for train set: 0.981429
classification accuracy for test set: 0.972750

Naive Bayes

classification accuracy for train set: 0.774143
classification accuracy for test set: 0.764250

Summary

Gaussian and K-NN performed similarly good, Naive Bayes performed worst which is as expected. Since we converted real valued features into binary features for Naive Bayes.