Q1

1. $$P(y=k|x,\mu,\sigma) = \frac{P(x,\mu,\sigma|y=k)\cdot P(y=k)}{P(x,\mu,\sigma)}$$

$$= \frac{P(x|\mu,\sigma,y=k)\,P(\mu,\sigma|y=k)\cdot P(y=k)}{P(x,\mu,\sigma)}$$

$$= \frac{P(x|\mu,\sigma,y=k)\cdot P(y=k)\cdot P(\mu,\sigma|y=k)}{P(x|\mu,\sigma)\cdot P(\mu,\sigma)}$$

$$= \frac{P(x|\mu,\sigma,y=k)\cdot P(y=k)}{\sum_{k=1}^{K} P(x|\mu,\sigma,y=k)\cdot P(y=k)}$$

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

2. $$-\log p(y^{(1)},x^{(1)},y^{(2)},x^{(2)},\dots,y^{(N)},x^{(N)}|\theta)$$

$$= -\log \prod_{j=1}^{N} (y^{(j)},x^{(j)}|\theta^{(j)})$$

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

$$= -\sum_{j=1}^{N}\left[\log \alpha_k^{(j)} - \frac{1}{2}\sum_{i=1}^{D}\log(2\pi\sigma_i^{(j)2}) - \sum_{i=1}^{D}\frac{1}{2\sigma_i^{(j)2}}(x_i^{(j)}-\mu_{ki}^{(j)})^2\right]$$

3. $$\frac{\partial l}{\partial \mu_{ki}} = \sum_{j=0}^{N}\mathbb{1}(y^{(j)}=k)\sum_{i=1}^{D}\frac{1}{\sigma_i^{(j)2}}(x_i^{(j)}-\mu_{ki}^{(j)})$$

$$\frac{\partial l}{\partial(\sigma_i^2)} = \sum_{j=0}^{N}\mathbb{1}(y^{(j)}=k)\cdot -\frac{1}{2}\cdot\sum_{i=1}^{D}\left[\frac{(x_i^{(j)}-\mu_{ki}^{(j)})^2}{z^2}-\frac{1}{z}\right], \text{ where } z=\sigma_i^{(j)2}$$

4. $\frac{\partial l}{\partial \mu_{ki}} = \sum_{j=0}^{N} \mathbb{1}(y^{(j)}=k) \sum_{i=1}^{D} \frac{1}{\sigma_i^{(j)2}} (x_i^{(j)} - \mu_{ki}^{(j)}) = 0$

$\Rightarrow \hat{\mu}_{ik} = \frac{1}{\sum_j \mathbb{1}(y^{(j)}=k)} \sum_j x_i^{(j)} \mathbb{1}(y^{(j)}=k)$

$\frac{\partial l}{\partial (\sigma_i^2)} = \frac{\partial l}{\partial z} = -\sum_{j=0}^{N} \mathbb{1}(y^{(j)}=k) \cdot \frac{1}{2} \cdot \sum_{i=1}^{D} \left[\frac{(x_i^{(j)} - \mu_{ki}^{(j)})^2 - z}{z^2} \right]$ where $z = \sigma_i^{(j)2}$

$\Rightarrow \hat{\sigma}_{ik}^2 = \frac{1}{\sum_j \mathbb{1}(y^{(j)}=k)} \sum_j (x_i^{(j)} - \mu_{ki}^{(j)})^2 \mathbb{1}(y^{(j)}=k)$

Q2.1

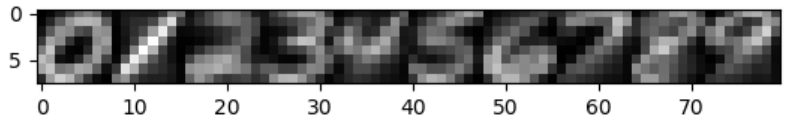
1. (a) $K=1$: Training accuracy is 1.0
Test accuracy is 0.86875

$K=15$: Training accuracy is 0.861
Test accuracy is 0.85625

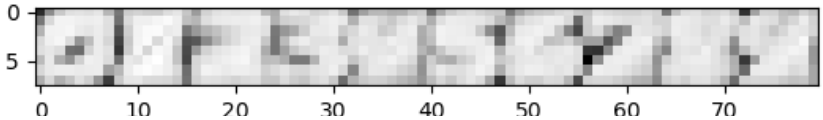
2. I use the ~~max~~ method of decreasing K by 1 until breaking the tie. This is a simple but effecient method, since a tie is impossible when $K=1$. This method can pursee the largest K without tie.

3. The optimal K I got is 3
Average accuracy across folds is 0.86514
Train classification accuracy is 0.88657
Test accuracy is 0.86875

2.0
plotting the original figure of mean



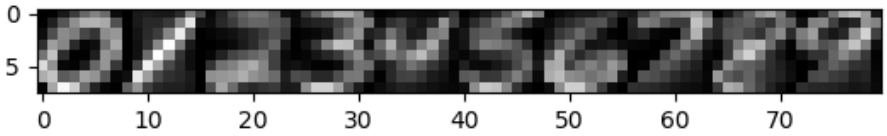
2.2
1.



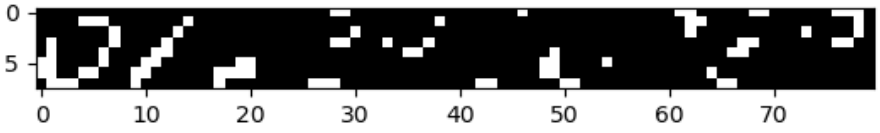
2. The average conditional likelihood of train data is -0.124624436669
   The average conditional likelihood of test data is -0.196673203255

3. The accuracy of the model on train data is 0.9814285714285714
   The accuracy of the model on test data is 0.97275

2.3
3.



4.



5. Since we use $P(y = C|\mathbf{x}) \propto P(y = C)P(\mathbf{x}|y = C)$,we ignore the P(x). The calculated conditional likelihood is smaller than the true value of P(y = c |x).

The average conditional likelihood of train data is -30.7898021047
The average conditional likelihood of test data is -30.7369008592

6. The accuracy of the model on train data is 0.7741428571428571
The accuracy of the model on test data is 0.76425

2.4 Model Comparison

Algorithms	KNN(K=3)	Conditional Gaussian	Naïve Bayes
Accuracy of train	0.98657	0.98143	0.77414
Accuracy of test	0.96975	0.97275	0.76425

From the table above, we see Naïve Bayes perform the worst in the analytics, KNN and Conditional Gaussian almost have the same good performance while Conditional Gaussian has a better performance on the prediction of the dataset. In our problem, the features are not purely independent, which cause Naïve Bayes to achieve just 77% accuracy in this problem. This matches my expectation!