CSC311H1 Assignmnet 4

Xiaoyu Zhou

September 10, 2020

1. (a) $p(y=k|x,\mu,\sigma)$
$= \frac{p(y=k,x,\mu,\sigma)}{p(x,\mu,\sigma)}$
$= \frac{p(x|y=k,\mu,\sigma)p(y=k|\mu,\sigma)}{p(x|\mu,\sigma)}$ (By using Bayes Rule)
$= \frac{p(x|y=k,\mu,\sigma)p(y=k|\mu,\sigma)}{\sum_{i=1}^{k} p(x|y=i,\mu,\sigma)p(y=i|\mu,\sigma)}$ (By law of total probability)
$= \frac{p(x|y=k,\mu,\sigma)p(y=k)}{\sum_{i=1}^{k} p(x|y=i,\mu,\sigma)p(y=i)}$
$= \frac{\left(\left(\prod_{i=1}^{D} 2\pi\sigma_i^2\right)^{-\frac{1}{2}} exp\left\{-\sum_{i=1}^{D} \frac{1}{2\sigma_i^2}(x_i-\mu_{ki})^2\right\}\right)a_k}{\sum_{j=1}^{k}\left(\left(\prod_{i=1}^{D} 2\pi\sigma_i^2\right)^{-\frac{1}{2}} exp\left\{-\sum_{i=1}^{D} \frac{1}{2\sigma_i^2}(x_i-\mu_{ki})^2\right\}\right)a_j}$ (By the definition from the question)

(b) $l(\theta;D)$
$= -\log p(y^{(1)},x^{(1)},y^{(2)},x^{(2)},...,y^{(N)},x^{(N)}|\theta)$
$= -\log \prod_{i=1}^{N} p(y^{(i)},x^{(i)}|\theta)$ (Since the data are iid)
$= -\log \prod_{i=1}^{N} \frac{p(y^{(i)},x^{(i)},\theta)}{p(\theta)}$
$= -\log \prod_{i=1}^{N} \frac{p(x^{(i)}|y^{(i)},\theta)p(y^{(i)}|\theta)p(\theta)}{p(\theta)}$
$= -\log \prod_{i=1}^{N} p(x^{(i)}|y^{(i)},\theta)p(y^{(i)}|\theta)$
$= -\log \prod_{i=1}^{N}\left[\left(\prod_{j=1}^{D} 2\pi\sigma_j^2\right)^{-\frac{1}{2}} \cdot exp\left\{-\sum_{j=1}^{D}\frac{1}{2\sigma_j^2}(x_j^{(i)}-\mu_{y^{(i)}j})^2\right\} \cdot \prod_{n=1}^{K}\mathbb{I}(y^{(i)}=n)a_n\right]$
$= -\log\left[\left(\prod_{j=1}^{D} 2\pi\sigma_j^2\right)^{-\frac{N}{2}} \cdot \prod_{i=1}^{N}\prod_{n=1}^{K}\mathbb{I}(y^{(i)}=n)a_n \cdot \prod_{i=1}^{N} exp\left\{-\sum_{j=1}^{D}\frac{1}{2\sigma_j^2}(x_j^{(i)}-\mu_{y^{(i)}j})^2\right\}\right]$
$= -\log\left[\left(\prod_{j=1}^{D} 2\pi\sigma_j^2\right)^{-\frac{N}{2}} \cdot \prod_{i=1}^{N}\prod_{n=1}^{K}\mathbb{I}(y^{(i)}=n)a_n \cdot exp\left\{-\sum_{i=1}^{N}\sum_{j=1}^{D}\frac{1}{2\sigma_j^2}(x_j^{(i)}-\mu_{y^{(i)}j})^2\right\}\right]$
$= -\log\left[\left(\prod_{j=1}^{D} 2\pi\sigma_j^2\right)^{-\frac{N}{2}}\right] - \log\left[\prod_{i=1}^{N}\prod_{n=1}^{K}\mathbb{I}(y^{(i)}=n)a_n\right] - \log\left[exp\left\{-\sum_{i=1}^{N}\sum_{j=1}^{D}\frac{1}{2\sigma_j^2}(x_j^{(i)}-\mu_{y^{(i)}j})^2\right\}\right]$

$= \frac{N}{2}\log\left[\prod_{j=1}^{D} 2\pi\sigma_j^2\right] - \log\left[\prod_{i=1}^{N}\prod_{n=1}^{K}\mathbb{I}(y^{(i)} = n)a_n\right] + \sum_{i=1}^{N}\sum_{j=1}^{D}\frac{1}{2\sigma_j^2}(x_j^{(i)} - \mu_{y^{(i)}j})^2$

$= \frac{N}{2}\sum_{j=1}^{D}\log\left[2\pi\sigma_j^2\right] - \sum_{i=1}^{N}\sum_{n=1}^{K}\mathbb{I}(y^{(i)} = n)\log a_n + \sum_{i=1}^{N}\sum_{j=1}^{D}\frac{1}{2\sigma_j^2}(x_j^{(i)} - \mu_{y^{(i)}j})^2$

(c) From previous part, the μ_{ki} that in the question is the μ_{kj} in my solution. Same with that, σ_i^2 is equal to σ_j^2
Also, when $y^{(i)} = k$, $\mu_{y^{(i)}j}$ is μ_{kj}.
Find partial derivatives of the likelihood with respect to μ_{kj}:

$\frac{\partial l}{\partial \mu_{kj}}$

$= \frac{\partial}{\partial \mu_{kj}}\left[\frac{N}{2}\sum_{j=1}^{D}\log\left[2\pi\sigma_j^2\right] - \sum_{i=1}^{N}\sum_{n=1}^{K}\mathbb{I}(y^{(i)} = n)\log a_n + \sum_{i=1}^{N}\sum_{j=1}^{D}\frac{1}{2\sigma_j^2}(x_j^{(i)} - \mu_{kj})^2\right]$

$= \frac{\partial}{\partial \mu_{kj}}\left(\sum_{i=1}^{N}\sum_{j=1}^{D}\frac{1}{2\sigma_j^2}(x_j^{(i)} - \mu_{kj})^2\right)$

$= \sum_{i=1}^{N}\frac{1}{2\sigma_j^2}(x_j^{(i)} - \mu_{kj})\cdot(-2)$

$= -\frac{1}{\sigma_j^2}\sum_{i=1}^{N}(x_j^{(i)} - \mu_{kj})$

Find partial derivatives of the likelihood with respect to σ_j^2:

$\frac{\partial l}{\partial \sigma_j^2}$

$= \frac{\partial}{\partial \sigma_j^2}\left[\frac{N}{2}\sum_{j=1}^{D}\log\left[2\pi\sigma_j^2\right] - \sum_{i=1}^{N}\sum_{n=1}^{K}\mathbb{I}(y^{(i)} = n)\log a_n + \sum_{i=1}^{N}\sum_{j=1}^{D}\frac{1}{2\sigma_j^2}(x_j^{(i)} - \mu_{kj})^2\right]$

$= \frac{\partial}{\partial \sigma_j^2}\left[\frac{N}{2}\sum_{j=1}^{D}\log\left[2\pi\sigma_j^2\right] + \sum_{i=1}^{N}\sum_{j=1}^{D}\frac{1}{2\sigma_j^2}(x_j^{(i)} - \mu_{kj})^2\right]$

$= \frac{N}{2}\sum_{j=1}^{D}\left(\frac{1}{2\pi\sigma_j^2}\cdot 2\pi\right) + \sum_{i=1}^{N}\sum_{j=1}^{D}\frac{1}{(2\sigma_j^2)^2}\cdot(-2)\cdot(x_j^{(i)} - \mu_{kj})^2$

$= \frac{N}{2\sigma_j^2} - \sum_{i=1}^{N}\frac{1}{2\sigma_j^4}(x_j^{(i)} - \mu_{kj})^2$

$= \frac{N}{2\sigma_j^2} - \frac{1}{2\sigma_j^4}\sum_{i=1}^{N}(x_j^{(i)} - \mu_{kj})^2$

(d) Find the maximum likelihood estimates for μ:
Let $\frac{\partial l}{\partial \mu_{kj}} = 0$.

$-\frac{1}{\sigma_j^2}\sum_{i=1}^{N}(x_j^{(i)} - \mu_{kj}) = 0$

$\sum_{i=1}^{N}(x_j^{(i)} - \mu_{kj}) = 0$

$\sum_{i=1}^{N}x_j^{(i)} = \sum_{i=1}^{N}\mu_{kj}$

$\mu_{kjMLE} = \frac{\sum_{i=1}^{N}x_j^{(i)}}{N}$

Find the maximum likelihood estimates for σ_j^2:

Let $\frac{\partial l}{\partial \sigma_j^2} = 0$.

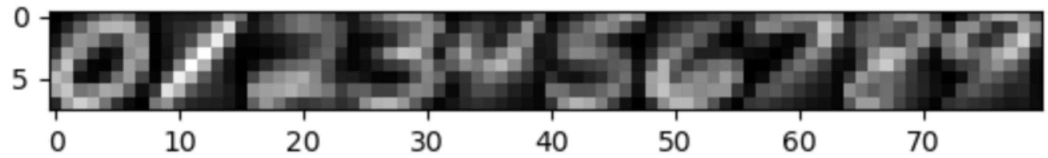
$\frac{N}{2\sigma_j^2} - \frac{1}{2\sigma_j^4} \sum_{i=1}^{N} (x_j^{(i)} - \mu_{kj})^2 = 0$

$\frac{N}{2\sigma_j^2} = \frac{1}{2\sigma_j^4} \sum_{i=1}^{N} (x_j^{(i)} - \mu_{kj})^2$

$\sigma_j^2 \cdot N = \sum_{i=1}^{N} (x_j^{(i)} - \mu_{kj})^2$

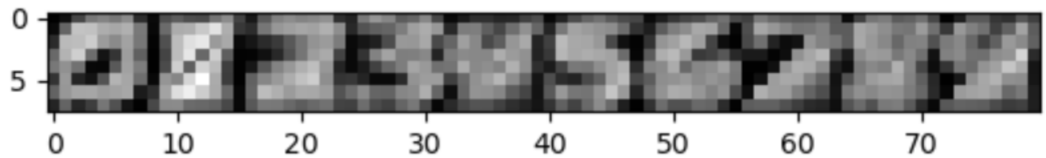
$\sigma_{jMLE}^2 = \frac{\sum_{i=1}^{N} (x_j^{(i)} - \mu_{kj})^2}{N}$

2. **2.0**



2.1 Conditional Gaussian Classifier Training

2.1.1



2.1.2

```
Average conditional log-likelihood for training data -0.1246244366686299
Average conditional log-likelihood for test data -0.19667320325525503
```

2.1.3

```
Conditional Gaussian classifier on training set has an accuracy:  0
 .9814285714285714
Conditional Gaussian classifier on test set has an accuracy:  0.97275
```

2.2 Naive Bayes Classifier Training

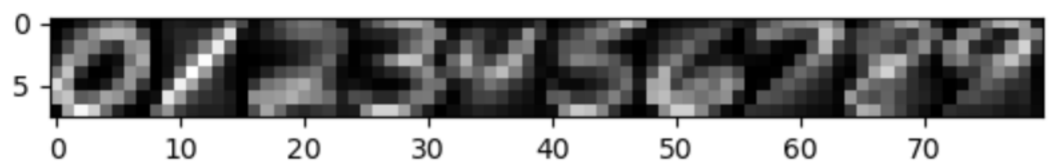
2.2.1

Please see the code.

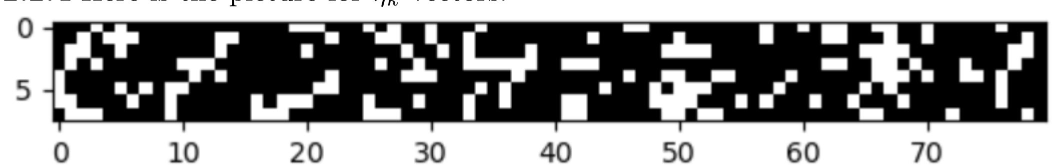
2.2.2

Please see the code.

2.2.3 Here is the picture for η_k vectors.



2.2.4 Here is the picture for η_k vectors.



2.2.5

```
Average conditional likelihood over the true training class labels:
 -0.9437538618002553
Average conditional likelihood over the true testing class labels: -0
 .987270433725358
```

2.2.6

```
The accuracy of naive bayes classifier on training set:  0.7741428571428571
The accuracy of naive bayes classifier on test set:  0.76425
```

2.3 Model Comparison

From the accuracy, we can see that Conditional Gaussian classifier performs better since its accuracy is over 0.97, while Naive Bayes classifier's accuracy is just around 0.77. The result fits my assumption since Naive Bayes classifier need the independence of the data. However, the relationship between every pixel is closed dependent to the actual digits. Thus, Naive Bayes classifier cannot perform well. At the same time, there are a large amount of accurate data for Conditional Gaussian classifier to fit, so it performs good.