CSC311 A3

Wilson Hsu

March 2020

Question 1

Part 1

Derive $p(y = k|x, \boldsymbol{\mu}, \boldsymbol{\sigma})$:

$$\begin{aligned}
p(y = k|x, \boldsymbol{\mu}, \boldsymbol{\sigma}) &= \frac{p(y = k, x, \boldsymbol{\mu}, \boldsymbol{\sigma})}{p(x, \boldsymbol{\mu}, \boldsymbol{\sigma})} \\
&= \frac{p(y = k, x, \boldsymbol{\mu}, \boldsymbol{\sigma})}{p(x|\boldsymbol{\mu}, \boldsymbol{\sigma})p(\boldsymbol{\mu}, \boldsymbol{\sigma})} \\
&= \frac{p(y = k, x|\boldsymbol{\mu}, \boldsymbol{\sigma})}{p(x|\boldsymbol{\mu}, \boldsymbol{\sigma})} \\
&= \frac{p(x|y = k, \boldsymbol{\mu}, \boldsymbol{\sigma})p(y = k)}{p(x|\boldsymbol{\mu}, \boldsymbol{\sigma})} \\
&= \frac{p(x|y = k, \boldsymbol{\mu}, \boldsymbol{\sigma})p(y = k)}{\sum_{j=1}^{k} p(x|y = j, \boldsymbol{\mu}, \boldsymbol{\sigma})p(y = j)} && \text{By Law of total probability}
\end{aligned}$$

Part 2

Derive $l(\boldsymbol{\theta}; \boldsymbol{x})$:

$$\begin{aligned}
l(\theta; \boldsymbol{x}) &= -log(p(y^{(1)}, \boldsymbol{x}^{(1)}...y^{(N)}, \boldsymbol{x}^{(N)}|\boldsymbol{\theta})) \\
&= -\sum_{i=1}^{N} log(p(y^{(i)}, \boldsymbol{x}^{(i)}|\boldsymbol{\theta})) && \text{Since the data are iid.} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{k} \mathbb{I}\{y^{(i)} = j\} log(p(y^{(i)} = j)p(\boldsymbol{x}^{(i)}|y^{(i)} = j, \boldsymbol{\theta})) && \text{Which are Equation (1) and (2)} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{k} \mathbb{I}\{y^{(i)} = j\} log(p(y^{(i)} = j)) - \sum_{i=1}^{N}\sum_{j=1}^{k} \mathbb{I}\{y^{(i)} = j\} log(p(\boldsymbol{x}^{(i)}|y^{(i)} = j, \boldsymbol{\theta}))
\end{aligned}$$

Part 3

Find $\frac{\partial l(\boldsymbol{\theta};\boldsymbol{x})}{\partial \mu_{ki}}$:

Only need to care about the terms where $y^{(j)} = k$ for the k that we are interested in.

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\theta};\boldsymbol{x})}{\partial \mu_{ki}} &= -\frac{\partial \sum_{j=1}^{N}\sum_{m=1}^{K}\mathbb{I}\{y^{(j)}=m\}log(p(\boldsymbol{x}^{(j)}|y^{(j)}=m,\boldsymbol{\theta}))}{\partial \mu_{ki}} && \text{Since partial derivative of first term is zero}\\
&= -\frac{\partial \sum_{j=1}^{N}\sum_{m=1}^{K}\mathbb{I}\{y^{(j)}=m\}log((\prod_{p=1}^{D}2\pi\sigma_p^2)^{-\frac{1}{2}})exp\{-\sum_{p=1}^{D}\frac{1}{2\sigma_p^2}(x_p^{(j)}-\mu_{mp})^2\})}{\partial \mu_{ki}} && \text{By Equation (1)}\\
&= -\frac{\partial(\sum_{j=1}^{N}\sum_{m=1}^{K}\mathbb{I}\{y^{(j)}=m\}(log((\prod_{p=1}^{D}2\pi\sigma_p^2)^{-\frac{1}{2}})-\sum_{p=1}^{D}\frac{1}{2\sigma_p^2}(x_p^{(j)}-\mu_{mp})^2))}{\partial \mu_{ki}}\\
&= \frac{\partial(\frac{N}{2}log((\prod_{p=1}^{D}2\pi\sigma_p^2))+\sum_{j=1}^{N}\sum_{p=1}^{D}\sum_{m=1}^{K}\mathbb{I}\{y^{(j)}=m\}\frac{1}{2\sigma_p^2}(x_p^{(j)}-\mu_{mp})^2)}{\partial \mu_{ki}}\\
&= \frac{\partial \sum_{j=1}^{N}\sum_{p=1}^{D}\sum_{m=1}^{K}\mathbb{I}\{y^{(j)}=m\}\frac{1}{2\sigma_p^2}(x_p^{(j)}-\mu_{mp})^2}{\partial \mu_{ki}} && \text{Since partial derivative of first term is zero}\\
&= \sum_{j=1}^{N}\frac{\mathbb{I}\{y^{(j)}=k\}(x_i^{(j)}-\mu_{ki})}{\sigma_i^2} && \text{Terms where } m\neq k \text{ or } p\neq i \text{ is constant w.r.t } \mu_{ki}
\end{aligned}$$

Find $\frac{\partial l(\boldsymbol{\theta};\boldsymbol{x})}{\partial \sigma_i^2}$:

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\theta};\boldsymbol{x})}{\partial \sigma_i^2} &= \frac{\partial(\frac{N}{2}log((\prod_{p=1}^{D}2\pi\sigma_p^2))+\sum_{j=1}^{N}\sum_{p=1}^{D}\sum_{m=1}^{K}\mathbb{I}\{y^{(j)}=m\}\frac{1}{2\sigma_p^2}(x_p^{(j)}-\mu_{mp})^2)}{\partial \sigma_i^2}\\
&= \frac{\partial(\frac{N}{2}\sum_{p=1}^{D}(log(2\pi\sigma_p^2))+\sum_{j=1}^{N}\sum_{p=1}^{D}\frac{1}{2\sigma_p^2}(x_p^{(j)}-\mu_{kp})^2)}{\partial \sigma_i^2}\\
&= \frac{N}{2\sigma_i^2}-\sum_{j=1}^{N}\frac{1}{2(\sigma_i^2)^2}(x_i^{(j)}-\mu_{ki})^2
\end{aligned}$$

Part 4

Find MLE for μ_{ki}:

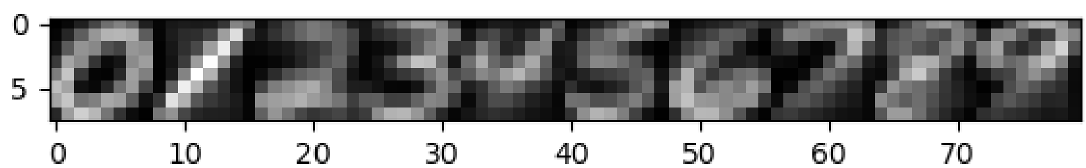
$$\begin{aligned}
0 &= \sum_{j=1}^{N}\frac{\mathbb{I}\{y^{(j)}=k\}(x_i^{(j)}-\mu_{ki})}{\sigma_i^2} && \text{Set } \frac{\partial l(\boldsymbol{\theta};\boldsymbol{x})}{\partial \mu_{ki}}=0\\
&\longrightarrow \sum_{j=1}^{N}\mathbb{I}\{y^{(j)}=k\}(x_i^{(j)}-\mu_{ki}) && \text{Equivalent problem as above}\\
&= \sum_{j=1}^{N}\mathbb{I}\{y^{(j)}=k\}(x_i^{(j)})-\sum_{j=1}^{N}\mathbb{I}\{y^{(j)}=k\}(\mu_{ki})\\
\sum_{j=1}^{N}\mathbb{I}\{y^{(j)}=k\}(x_i^{(j)}) &= \sum_{j=1}^{N}\mathbb{I}\{y^{(j)}=k\}(\mu_{ki})\\
\sum_{j=1}^{N}\mathbb{I}\{y^{(j)}=k\}(x_i^{(j)}) &= \mu_{ki}\sum_{j=1}^{N}\mathbb{I}\{y^{(j)}=k\}\\
\mu_{ki} &= \frac{\sum_{j=1}^{N}\mathbb{I}\{y^{(j)}=k\}(x_i^{(j)})}{\sum_{j=1}^{N}\mathbb{I}\{y^{(j)}=k\}}
\end{aligned}$$

Find MLE for σ_i^2:

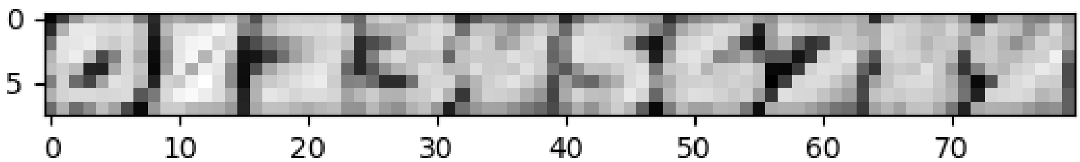
$$
\begin{aligned}
0 &= \frac{N}{2\sigma_i^2} - \sum_{j=1}^{N} \frac{1}{2(\sigma_i^2)^2}(x_i^{(j)} - \mu_{ki})^2 && \text{Set } \frac{\partial l(\boldsymbol{\theta};\boldsymbol{x})}{\partial \sigma_i^2} = 0 \\
\frac{N}{2\sigma_i^2} &= \frac{1}{2(\sigma_i^2)^2} \sum_{j=1}^{N}(x_i^{(j)} - \mu_{ki})^2 \\
\sigma_i^2 &= \frac{\sum_{j=1}^{N}(x_i^{(j)} - \mu_{ki})^2}{N}
\end{aligned}
$$

Question 2

Part 2.0



Part 2.1.1



Part 2.1.2

```
Avg conditional log likelihood (Train): -0.12462443666862995
Avg conditional log likelihood (Test): -0.1966732032552551
```

Part 2.1.3

```
Accuracy (Train): 0.9814285714285714
Accuracy (Test): 0.97275
```

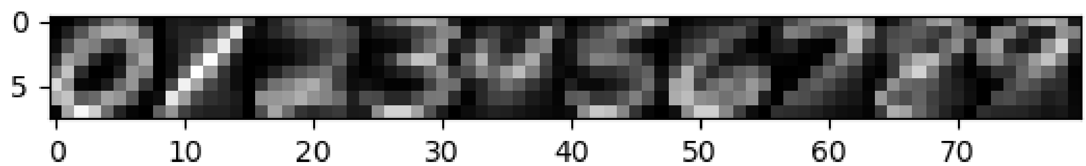
Part 2.2.1

Given in starter code

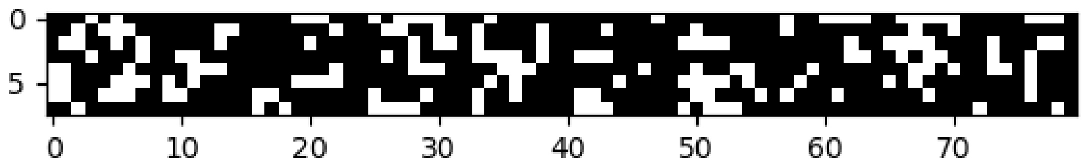
Part 2.2.2

In *q*2_2.py file

Part 2.2.3



Part 2.2.4



Part 2.2.5

```
Avg conditional log likelihood (Train): -0.9437538618002539
Avg conditional log likelihood (Test): -0.9872704337253584
```

Part 2.2.6

```
Accuracy (Train): 0.7741428571428571
Accuracy (Test): 0.76425
```

Part 2.3

Conditional Gaussian Classifier has accuracy of 0.98 on Training data and 0.97 on Test data, while Naive Bayes Classifier has accuracy of 0.77 on Training data and 0.76 on Test data. Conditional Gaussian Classifier performed the best out of 2 with the significantly higher accuracy and Naive Bayes Classifier performed the worst. This matched my expectation as Naive Bayes Classifier uses an assumption that all features are conditionally independent given class c, which is not the case as the relationship between features (pixels) in this data set is not independent.