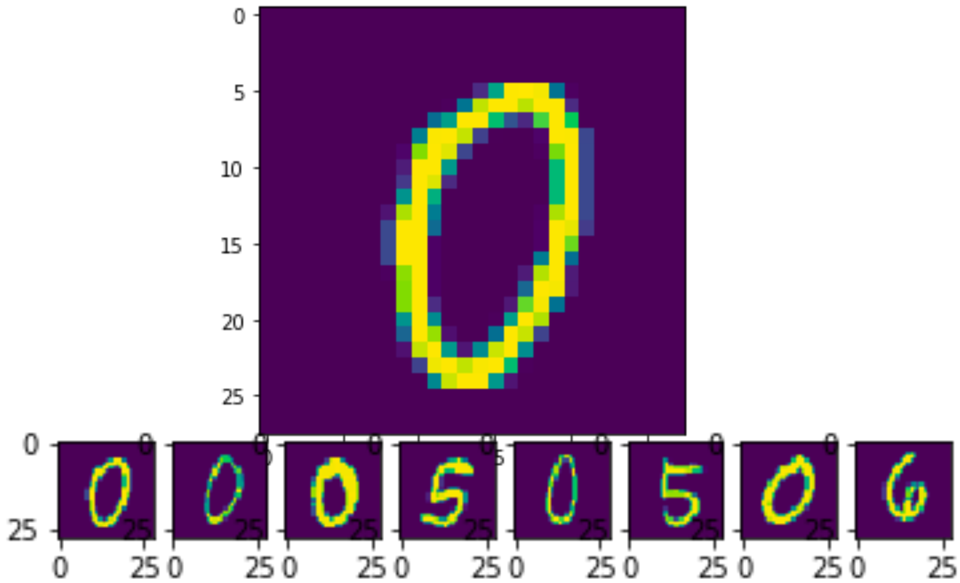
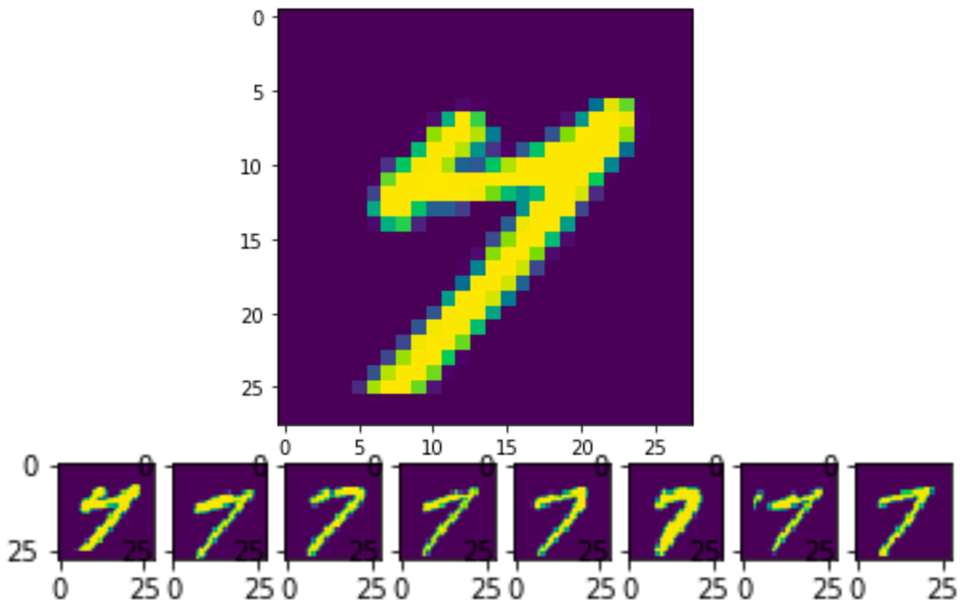
Machine Learning HW2

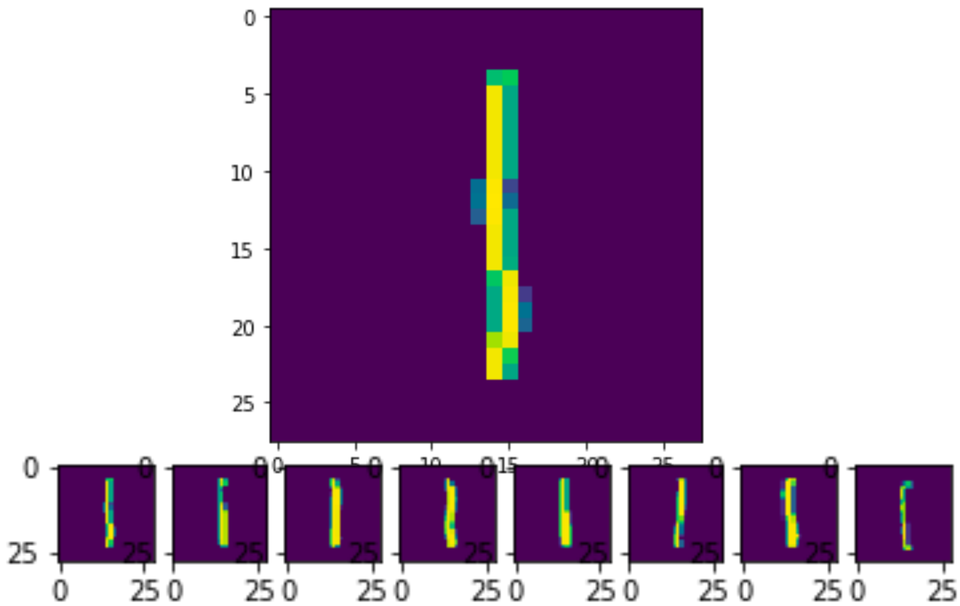
Fall 2020

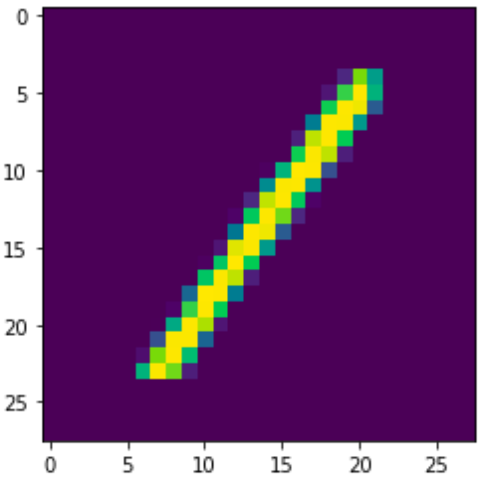
PoKang Chen

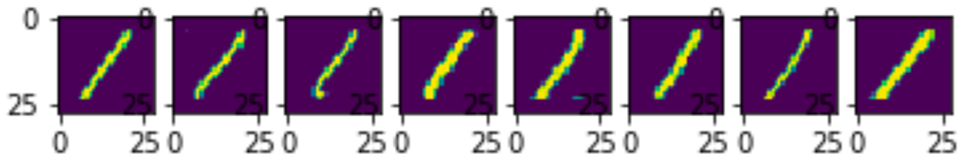
1.(a)

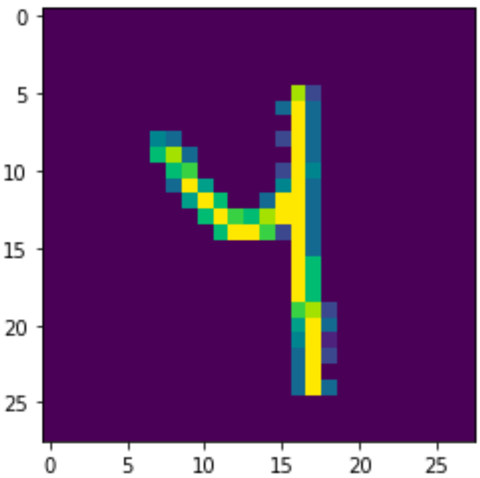


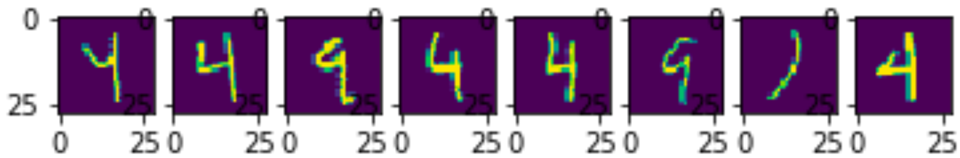












1.(b)

When k = 10, accuracy = 0.857

1.(c)

When having over 100,000 training examples, the computing time would be t oo much. This is one of the disadvantages of kNN search. Further,as the num ber k increases, the accuracy would then decrease.

2.(a)

2.
(a) When there're different $\underline{w}$

$\underline{\text{log likelihood}}$:

$$\ell(w) = \sum_{i}^{N} \log \prod_{k=1}^{K-1} \left[\frac{e^{(w_k^T \phi(x^{(i)}))}}{1+\sum_{j=1}^{M} e(w_j^T \phi(x^{(i)}))} \right]^{I(y^{(i)}=k)} + \left[\log \frac{1}{1+\sum_{j=1}^{K-1} e^{w_j^T \phi(x)}} \right]^{I(y^{(i)}=K)}$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K-1} I\{y^{(i)}=k\} \left[w_k^T \phi(x) - \log\left(1+\sum_{j=1}^{M} e^{(w_j^T \phi(x^{(i)}))}\right) \right] - I\{y^{(i)}=K\} \left[\log\left(1+\sum_{j=1}^{K-1} e^{(w_j^T \phi(x))}\right) \right]$$

$\nabla_{w_m} \ell(w)$

$$= \sum_{i=1}^{N} I\{y^{(i)}=m\} \left[\phi(x^{(i)}) - \underbrace{\frac{1}{1+\sum_{j=1}^{K} e^{(w_j^T \phi(x))}} e^{w_j^T \phi(x^{(i)})}}_{p(y^{(i)}=m \mid w, \phi(x^{(i)}))} \phi(x^{(i)}) \right] + I\{y^{(i)} \neq m\} \left[\underbrace{\frac{1}{1+\sum_{j=1}^{K} e^{w_j^T x^{(i)}}} e^{w_k^T \phi(x^{(i)})} \phi(x^{(i)})}_{p(y^{(i)}=m \mid w, \phi(x^{(i)}))} \right]$$

$1 \sim K-1$

For kth category, only $\nabla \sum_{j=1}^{K-1} e^{w_j^T x^{(i)}}$ is nonzero, which is $e^{w_k^T x^{(i)}}$.

$$\therefore = \sum_{i=1}^{N} \left[\phi(x^{(i)}) \left(I\{y^{(i)}=m\} - p(y^{(i)}=m \mid w, \phi(x^{(i)})) \right) \right]$$

2.(b)

Train = 0.97
Test = 0.92

3.(a)

3. (a)

Since $p(y|x) = \frac{p(x|y)\cdot p(y)}{p(x)} = \frac{p(x|y)\cdot p(y)}{p(x|y=0)\cdot p(y=0) + p(x|y=1)p(y=1)}$

$A = -\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)$

$B = -\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)$

$C = \frac{1}{(2\pi)^{M/2}|\Sigma|^{1/2}}$

$\therefore$ When $y=0$,

$$p(y=0|x) = \frac{p(x|y)\,p(y=0)}{\Sigma p(x|y')\,p(y')} = \frac{e^A\phi}{e^A\phi + e^B(1-\phi)} = \frac{1}{1+\frac{1-\phi}{\phi}e^{(B-A)}}$$

$$= \frac{1}{1+e^{(B-A+\ln\frac{1-\phi}{\phi})}} = \frac{1}{1+e^{-(A-B+\ln\frac{\phi}{1-\phi})}}$$

When $y=1$,

$$p(y=1|x) = \frac{p(x|y)\,p(y=1)}{\Sigma p(x|y)\,p(y)} = \frac{1}{\frac{\phi}{1-\phi}e^{(A-B)}+1} = \frac{1}{1+e^{-(A-B+\ln\frac{\phi}{1-\phi})}}$$

Same.

To show $p(y|x)$ is in form of a logistic function: $\boxed{p(y|x) = \frac{1}{1+e^{-(w^Tx)}}}$

$\Rightarrow A-B+\ln\frac{\phi}{1-\phi}$

$= -\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0) - \frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + \ln\frac{\phi}{1-\phi}$

$= -\frac{1}{2}(x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu_0 - \mu_0^T\Sigma^{-1}x + \mu_0^T\Sigma^{-1}\mu_0 + x^T\Sigma^{-1}x - x\Sigma^{-1}\mu_1 - \mu_1^T\Sigma^{-1}x + \mu_1^T\Sigma^{-1}\mu_1) + \ln\frac{\phi}{1-\phi}$

$= \underbrace{(\mu_1-\mu_0)^T\Sigma^{-1}}_{W^T}x + \underbrace{\frac{1}{2}(\mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}\mu_1) + \ln\frac{\phi}{1-\phi}}_{w_0}$

3.(b)(c)

(b)(c) The log likelihood:

$$\ell(\phi,\mu_0,\mu_1,\Sigma) = \log \prod_{i=1}^{N} p(x^{(i)}, y^{(i)}; \phi,\mu_0,\mu_1,\Sigma) = \log \prod_{i=1}^{N} p(x^{(i)}|y^{(i)}; \phi,\mu_0,\mu_1,\Sigma) \times p(y^{(i)};\phi)$$

$$= \log \prod_{i=1}^{N}\left[\ell\{y^{(i)}=0\} \frac{1}{(2\pi\sigma^2)^{1/2}} e^{\left(-\frac{(x^{(i)}-\mu_0)^2}{2\sigma^2}\right)} \phi + \ell\{y^{(i)}=1\} \frac{1}{(2\pi\sigma^2)^{1/2}} e^{\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right)} (1-\phi)\right]$$

$$\ell = \log\left[\left(\prod_{i=1}^{N} \frac{1}{(2\pi\sigma^2)^{1/2}} e^{\left(-\frac{(x^{(i)}-\mu_0)^2}{2\sigma^2}\right)}\phi\right)\left(\prod_{j=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(x^{(i)}-\mu_1)^2}{2\sigma^2}\right)}(1-\phi)\right)\right]$$

$N_0 + N_1 = N'$

$N_0 = \sum_{i=1}^{N} \ell\{y^{(i)}=0\}$

$N_1 = \sum_{i=1}^{N} \ell\{y^{(i)}=1\}$

$$= \sum_{i=1}^{N'} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \sum_{i=1}^{N_0}\left(-\frac{(x^{(i)}-\mu_0)^2}{2\sigma^2}\right) + \sum_{i=1}^{N_0}\log\phi + \sum_{i=1}^{N_1}\left(-\frac{(x^{(i)}-\mu_1)^2}{2\sigma^2}\right) + \sum_{i=1}^{N_1}\log(1-\phi)$$

$$= -\frac{N'}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{N_0}\frac{(x^{(i)}-\mu_0)^2}{2\sigma^2} - \sum_{j=1}^{N_1}\frac{(x^{(i)}-\mu_1)^2}{2\sigma^2} + N_0\log\phi + N_1\log(1-\phi)$$

Maximize likelihood

$$\frac{\partial \ell}{\partial \phi} = \frac{N_0}{\phi} - \frac{N_1}{1-\phi} = 0 \Rightarrow \phi = \frac{N_0}{N_0+N_1} = \frac{1}{N}\sum_{i=1}^{N}\ell\{y^{(i)}=1\} \quad \#$$

$$\frac{\partial \ell}{\partial \mu_0} = -\frac{1}{2\sigma^2}\sum_{i=1}^{N_0} 2(x^{(i)}-\mu_0)\cdot(-1) = \frac{1}{\sigma^2}\left(\sum_{i=1}^{N}\ell\{y^{(i)}=1\}x^{(i)} - \sum_{i=1}^{N}\ell\{y^{(i)}=1\}\mu_0\right)$$

$$\Rightarrow \frac{\partial \ell}{\partial \mu_0} = 0 \Rightarrow \mu_0 = \frac{\sum_{i=1}^{N}\ell\{y^{(i)}=0\}x^{(i)}}{\sum_{i=1}^{N}\ell\{y^{(i)}=0\}} \quad \#$$

$$\frac{\partial \ell}{\partial \sigma^2} = 0 \Rightarrow \frac{\partial \ell}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \sum_{i=1}^{N_0}\frac{(x^{(i)}-\mu_0)^2}{2\sigma^4} + \sum_{i=1}^{N_1}\frac{(x^{(i)}-\mu_1)^2}{2\sigma^4} = 0.$$

$$\therefore \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x^{(i)}-\mu_{y^{(i)}})(x^{(i)}-\mu_{y^{(i)}})^T$$

$$\therefore \mu_1 = \frac{\sum_{i=1}^{N}\ell\{y^{(i)}=1\}x^{(i)}}{\sum_{i=1}^{N}\ell\{y^{(i)}=1\}} \quad \#$$

(b)(c)

$$\ell = -\frac{N\pi}{2}\log(2\pi) - \frac{N}{2}\log(|\Sigma|) - \sum_{i=1}^{N_0}\left(\frac{1}{2}(x^{(i)}-\mu_0)^T\Sigma^{-1}(x^{(i)}-\mu_1)\right) - \sum_{i=1}^{N_1}\left(\frac{1}{2}(x^{(i)}-\mu_1)^T\Sigma^{-1}(x^{(i)}-\mu_1)\right) + N_0\log\phi + N_1\log(1-\phi)$$

$$\therefore \nabla \ell = -\sum_{i=1}^{N_0}\frac{1}{2}\cdot 2\Sigma^{-1}(x^{(i)}-\mu_0)(-1) = \Sigma^{-1}\sum_{i=1}^{N_0}(x^{(i)}-\mu_0) = (\Sigma^{-1})\left(\sum_{i=1}^{N}\ell\{y^{(i)}=1\}x^{(i)} - \sum_{i=1}^{N}\ell\{y^{(i)}=1\}\mu_1\right)$$

$$\nabla \ell = 0 \qquad \mu_0 = \frac{\sum_{i=1}^{N}\ell\{y^{(i)}=0\}x^{(i)}}{\sum_{i=1}^{N}\ell\{y^{(i)}=0\}} \qquad \mu_1 = \frac{\sum_{i=1}^{N}\ell\{y^{(i)}=1\}x^{(i)}}{\sum_{i=1}^{N}\ell\{y^{(i)}=1\}}.$$

$$\nabla_{\Sigma^{-1}}\left(-\frac{N}{2}\log|\Sigma|\right) = \nabla_{\Sigma^{-1}}\left(-\frac{N}{2}\log\left(\frac{1}{|\Sigma^{-1}|}\right)\right) = \nabla_{\Sigma^{-1}}\left(\frac{N}{2}\log(|\Sigma^{-1}|)\right) = \frac{N}{2}(\Sigma^{-1})^{-1} = \frac{N}{2}\Sigma$$

$$\nabla_{\Sigma^{-1}}\left(-\sum_{i=1}^{N_0}\left(-\frac{1}{2}(x^{(i)}-\mu_0)^T\Sigma^{-1}(x^{(i)}-\mu_0)\right) - \sum_{i=1}^{N_1}\left(-\frac{1}{2}(x^{(i)}-\mu_1)^T\Sigma^{-1}(x^{(i)}-\mu_1)\right)\right) = -\frac{1}{2}\sum_{i=1}^{N}(x^{(i)}-\mu_{y^{(i)}})(x^{(i)}-\mu_{y^{(i)}})^T$$

$$\frac{N}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{N}(x^{(i)}-\mu_{y^{(i)}})(x^{(i)}-\mu_{y^{(i)}})^T = 0$$

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}(x^{(i)}-\mu_{y^{(i)}})(x^{(i)}-\mu_{y^{(i)}})^T$$

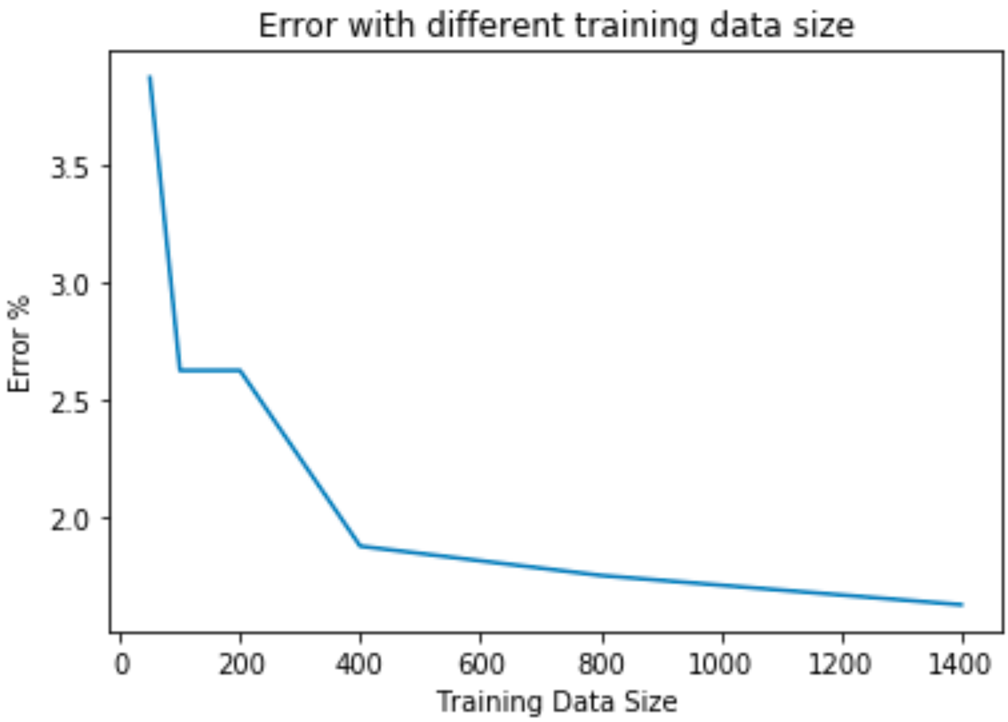
4(a)

Error: 1.6250%

4(b)

['httpaddr', 'spam', 'unsubscrib', 'ebai', 'valet', 'diploma',
'dvd', 'websit', 'click', 'lowest']

4(c)



Error: 1.6250%
Error: 3.8750%
Error: 2.6250%
Error: 2.6250%
Error: 1.8750%
Error: 1.7500%
Error: 1.6250%